

Уральский государственный университет им. А. М. Горького

Вл. Д. Мазуров

**МАТЕМАТИЧЕСКИЕ МЕТОДЫ
РАСПОЗНАВАНИЯ ОБРАЗОВ**

Учебное пособие

Екатеринбург
2010

УДК 004.93:51
М 139

Вл. Д. Мазуров. Математические методы распознавания образов. Уч. пособ. 2-е изд., доп. и перераб. – Екатеринбург: Изд-во Урал. ун-та, 2010. – 101 с.

В учебном пособии излагаются основные математические модели и методы решения задач распознавания образов

© Мазуров Вл. Д., 2010

ВВЕДЕНИЕ

Задачи распознавания образов связаны с классификацией объектов, которая подчинена некоторой цели. Существуют различные направления формализации этих задач, приводящие к множеству математических постановок. Одно из направлений использует теорию и методы решения систем линейных неравенств. Именно это направление, основанное на алгебраической теории линейных неравенств, развитой С. Н. Черниковым, представлено в данном пособии.

Распознавание образов – относительно новая область математической кибернетики, но она интенсивно развивается, и уже сейчас имеются многочисленные важные приложения методов распознавания буквально во всех сферах науки и техники, в том числе в экономике, технике, медицине, физике, биологии, социологии и так далее. Распознавание образов является основой искусственного интеллекта.

Общая структура курса, излагаемого в пособии, такова: рассматриваются математические модели классификации; вычислительные методы распознавания образов, главным образом, итерационные методы решения систем линейных неравенств; приложения, в том числе в математическом программировании, экономике, технике, биологии и медицине.

Следует подчеркнуть возрастающую роль итерационных методов распознавания образов и математического программирования (в частности, методов фейеровского типа, развитых в работах И. И. Ерёмина). Это связано с усложнением практических задач, ростом их нестационарности, увеличением размерности.

Что касается специфики задач распознавания образов по сравнению с традиционными для теории линейных неравенств приложениями, то она состоит, в частности, в

следующем. При решении задач дискриминантного анализа необходимо решать вопрос выбора класса разделяющих функций: либо рассматривать всё более широкие классы, чтобы получить в конце концов разрешимую систему линейных неравенств, либо создавать конструкции (в том числе комитетных) для несовместных систем неравенств.

Приспособление методов к этой специфике осложняется тем, что разделяемые конечные множества растут с течением времени.

Есть и другие проблемы теории распознавания образов:

- нахождение условий разделимости конечных множеств функциями заданного класса;
- установление сходимости вычислительных методов (в том числе методов оптимизации), включающих блоки распознавания образов;
- исследование эффективности вычислительных алгоритмов распознавания;
- получение оценок надежности прогноза с помощью методов классификации.

Глава 1.

ОСНОВНЫЕ МАТЕМАТИЧЕСКИЕ МОДЕЛИ РАСПОЗНАВАНИЯ ОБРАЗОВ

§ 1. Содержательный смысл задачи распознавания образов

1.1. Общее определение

Задача распознавания образов вообще есть задача описания и классификации (узнавания, диагностики) объектов той или иной природы, это задача выработки понятия о классе объектов. Другими словами, одна из задач распознавания образов состоит в целесообразном разбиении заданной совокупности объектов на однородные классы. Объекты, вошедшие в один класс, должны быть «похожи» друг на друга или близки друг к другу по некоторому критерию «похожести» или близости. Объекты же из разных классов должны быть достаточно различными, далёкими друг от друга.

В соответствии с этим образ есть множество всех объектов, сходных друг с другом в некотором фиксированном отношении. Распознать образ объекта – значит указать, к какому образу (классу похожих объектов) он относится.

1.2. Примеры

1. *Узнавание объектов.* Узнавание человеком объектов состоит в том, что он относит их к классу сходных объектов. Множество всех предметов, например, человек делит на классы и каждому классу присваивает имя: «столы»,

«птицы», «буква «а»» и т.п. Человек распознает как зрительные, так и звуковые сигналы.

2. *Медицинская диагностика.* В задаче медицинской диагностики образом является множество всех людей, страдающих определенным заболеванием. На материале обследованных с верифицированным диагнозом требуется выработать решающее правило, позволяющее отнести очередного пациента к тому или иному образу.

3. *Техническая диагностика.* Установление типа неисправности сложной технической системы является непростой задачей. Требуется найти решающее правило, позволяющее по набору тестов и косвенных признаков найти вид неисправности.

4. *Классификация в биологии.* Систематика является самостоятельным разделом биологии, она непрерывно развивается. Иерархическое разбиение живых объектов на классы позволяет структурировать многообразие живого и дает возможность определять место и роль каждого нового факта.

5. *Классификация в социологии.* Одна из задач социологического исследования той или иной группы населения: разделение её на однородные в каком-либо отношении подгруппы и выделение типопредставителей в каждой подгруппе.

6. *Типология систем.* Обобщение задачи предыдущего пункта: некоторый класс систем (сложных объектов, технико-экономических, природных и других) требуется разбить на однородные группы с выделением типопредставителей в каждой группе.

7. *Классификация и диагностика ситуаций.* Человек научается распознавать тип ситуации, в которой он находится, и принимать соответствующее, заранее заготовленное для этого типа решение. Конкретным примером является распознавание производственных ситуаций диспетчером, управляющим некоторым технологическим процессом.

8. *Разбиение пространства.* Пусть M и N – выпуклые телесные непересекающиеся многогранные множества в пространстве R^n . Следующая задача имеет многочисленные важные применения, в частности, в математической экономике: найти полупространство $P \subset R^n$ такое, чтобы $P \supset M$, $P \cap N = \emptyset$. Это полупространство определяет разбиение R^n на два класса: P и $R^n \setminus P$.

1.3. Значение операции классификации

Классификация есть простейшая логическая операция, однако она играет фундаментальную роль в мышлении, поскольку является одной из основ более сложных процессов. В связи с этим формализация мыслительных процессов, проводимая в рамках проблемы «Искусственный интеллект», предусматривает прежде всего формализацию операций классификации. Этим объясняется широкая применимость методов распознавания образов.

Различные научные теории имеют одинаковые способы организации знания, такие, например, как систематизация объектов исследования (эмпирических или абстрактных). Эта систематизация может как принимать форму простой классификации, так и иметь более сложный вид – до построения хорошо организованной системы. Распознавание образов является одним из первоначальных этапов организации данных, средством типизации и систематики.

1.4 Необходимость математических методов классификации

Исследования в области распознавания образов, в конечном счете, направлены на создание программ, обучающихся классификации и диагностике. Почему необходимы такие программы?

Дело в том, что обычные интуитивные методы, которые применяет человек, оказываются неэффективными, когда речь идет о типологии современных сложных производственных или иных систем.

Человек хорошо распознает зрительные образы: предметы, изображения (рисунки, чертежи), а также легко распознает слуховые образы. Но человек испытывает сложности при распознавании большого количества числового материала. По числовым показателям приборов, по таблицам трудно принять оперативное решение, если число показателей достаточно велико.

Современная же наука вообще характеризуется непрерывным возрастанием сложности изучаемых систем ввиду постоянного усложнения техники и организации производства, расширения ассортимента и сложности выпускаемой продукции, а также необходимости более глубокого анализа сущности природных, технических и экономических явлений, при котором учитываются самые слабые взаимные влияния элементов систем. При прогнозировании ситуации и характеристик объектов с помощью математических моделей приходится учитывать большое число взаимосвязанных факторов в условиях, когда доступная исследователю информация может иметь самые разнообразные формы и быть различной по степени детерминированности.

Большие трудности связаны с моделированием нестационарных и плохо определенных процессов, расплывчатых целей, нечетких ограничений. В качестве

примеров систем, при прогнозировании развития которых возникают принципиально новые трудности, можно назвать следующие:

- большие системы экономики;
- сложные производственные системы;
- экологические системы;
- социологические системы.

Так, например, если и могут быть построены неидентифицированные модели для реальных экологических систем, то измерения числовых параметров (таких, как численность популяции) связаны с применением более тонкого аппарата, чем детерминистские вычисления; они должны проводиться на основе самонастраивающихся и самообучающихся программ. Эффективным средством учёта плохо формализуемых соотношений при моделировании сложных систем является распознавание образов (РО).

Кроме того, сейчас возникла необходимость заменить человека в оперативном управлении быстропротекающими производственными процессами, в условиях управления большими системами с большим числом факторов; создаются АСУ, и нужно автоматизировать операции распознавания объектов и ситуаций.

Существует еще один аспект, с точки зрения чистой теории, познания законов мышления.

1.5. Схема применения методов распознавания образов в прогнозировании

Прежде чем перейти к формальным моделям распознавания образов коснёмся еще раз вопроса области приложений методов классификации. Начнем с примера применения распознавания образов в прогнозировании.

Предположим, что мы научились (с помощью ЭВМ или других средств) распознавать тип ситуации, а следовательно,

и выбирать то или иное решение в некоторой конкретной области производства. Мы в состоянии также вписать варианты будущих ситуаций. Применяя распознавание образов, мы можем в этом случае сделать прогноз множества будущих решений.

Этот подход можно иллюстрировать следующим более конкретным примером. Допустим, что имеются данные об открытых к настоящему времени месторождениях полезных ископаемых, служащих сырьем для металлургического производства. Можно сформировать различные варианты смесей материалов, получаемых из различных месторождений, а также варианты их предварительной подготовки как металлургического сырья. Предположим также, что на основании накопленного опыта мы выработали метод распознавания технологических характеристик смесей веществ по параметрам компонентов смеси и по параметрам, характеризующим способ подготовки. Тогда мы можем составить прогноз сырьевой базы металлургии.

§ 2. Моделирование объекта классификации

2.1. Вектор состояния

Простейшая модель, применяемая в распознавании образов, есть вектор состояния. На основе этой модели строятся более сложные. Модель объекта, или вектор состояния объекта, имеет вид:

$$x = (\xi_1, \dots, \xi_n),$$

где ξ_i - вещественные числа, интерпретируемые как значения параметров или признаков объекта.

Мы будем считать, что $x = (\xi_1, \dots, \xi_n)$ есть вектор пространства R^n . И использовать операции сложения векторов, умножения вектора на скаляр, а также скалярного произведения двух векторов. Скалярное произведение векторов $x, y \in R^n$ будем означать через $\langle x, y \rangle$.

2.2. Примеры

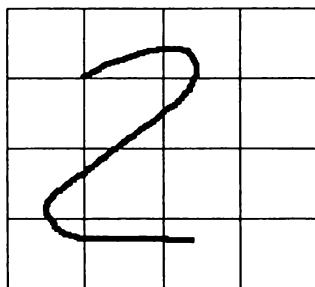
1. Объект – машиностроительное предприятие; ξ_1 – годовой объем выпускаемой продукции (в рублях); ξ_2 – число рабочих и служащих; ξ_3 – основные фонды; ξ_4 – количество поставщиков; ξ_5 – число потребителей продукции завода. Тогда модель объекта имеет вид:

$$x = (\xi_1, \dots, \xi_5) \in R^5.$$

2. Объект – человек. Признаки – вопросы анкеты в некотором социологическом обследовании. Ответ «да» кодируется числом 1, ответ «нет» - числом 0.

2.3. Модель изображения

Один из возможных способов предоставления изображений в виде n – мерного вектора таков. Изображение находится на плоском экране, который разбит на элементарные ячейки – прямоугольники. Эти элементы просматриваются строка за строкой слева направо и сверху вниз, и в качестве i -й координаты записывается 1, если в i -ю ячейку попал элемент изображения, иначе записывается 0 (см. рис. 1)



$x = (0, 1, 1, 0, 0, 1, 1, 0, 1, 1, 1, 0, 1, 1, 1, 0)$

Рис. 1.

§ 3. Модель дискриминантного анализа

3.1. Модель образа

Математической моделью образа (класса объектов) как некоторого множества всех объектов, сходных друг с другом в каком-то фиксированном отношении, является подмножество в пространстве $R^n: M \subset R^n$.

3.2. Пример

Объекты – люди. Они моделируются 2-мерными векторами: $x = (\xi_1, \xi_2)$, где ξ_1 – температура тела, ξ_2 – количество ударов сердца в минуту. Будем условно считать человека здоровым, если его температура колеблется в пределах от 35 до 37°C, а пульс – от 60 до 70 ударов в минуту. Понятно, что состояние человека не отражается полностью этими двумя показателями, пример чисто условный. При указанных предложениях образ «здоровые люди» имеет вид:

$$M = \{x = (\xi_1, \xi_2): 35 \leq \xi_1 \leq 37, 60 \leq \xi_2 \leq 70\}.$$

3.3. Идея обучения на прецедентах

В модели дискриминантного анализа реализуется следующая схема обучения на примерах. Пусть мы хотим обучить кого-нибудь отличать объекты одного рода (первого образа) от объектов другого рода (второго образа). Мы не пытаемся объяснить, в чем именно состоит отличие первых объектов от других, а просто показываем конечное число объектов того и другого вида. На основании этой информации у обучаемого должно сформироваться правило, позволяющее

любой объект относить либо к первому, либо ко второму образу.

3.4. Задача дискриминантного анализа для двух образов

Пусть имеется известное нам разбиение пространства R^n на два класса (образа): $R^n = A \cup B$, $A \cap B = \emptyset$. Предполагается, что известны лишь некоторые два подмножества: $A' \subset A$, $B' \subset B$. По этим подмножествам надо составить представление (разумеется, в общем случае лишь приближенное) о множествах A и B .

Делается это следующим образом. Пусть выбран некоторый класс функций

$$F \subset \{R^n \rightarrow R^1\},$$

с помощью которых будем отделять множество A' от множества B' . Требуется найти функцию $f \in F$, такую что

$$f(x) > 0 \ (\forall x \in A'), \ f(x) \leq 0 \ (\forall x \in B'). \quad (1)$$

При этом, на функцию f может ещё дополнительно накладываться некоторое условие качества разделения множеств (будем называть это условие условием нормализации).

Задача нахождения такой разделяющей или дискриминантной функции и есть задача дискриминантного анализа.

Если f - решение системы (1), удовлетворяющее условию нормализации, то мы получаем следующее решающее правило для прогноза принадлежности

произвольного элемента y к тому или иному образу:

$$y \in \begin{cases} A, & \text{если } f(y) > 0; \\ \text{иначе } B \end{cases}.$$

Следовательно, приближенной модель множества A будет $\{y: f(y) > 0\}$.

3.5. Пример

Рассмотрим условный пример. Обозначим через A' конечное множество видов технологически допустимого сырья в некотором производстве, через B' – конечное множество видов технологически недопустимого сырья. Каждый вид сырья пусть характеризуется двумя параметрами: ξ_1 – содержание первого компонента в единице сырья, ξ_2 – содержание второго, т.е. модель каждого вида сырья есть 2-мерный вектор $x = (\xi_1, \xi_2)$.

Пусть $A' = \{(1,3); (2,1); (1,1)\}; B' = \{(5,1); (5,2); (2,7)\}$, и F – класс аффинных функций, т.е. $f \in F \Leftrightarrow f(x) = \alpha_1 \xi_1 + \alpha_2 \xi_2 + \beta$. Составляем систему (1):

$$\begin{aligned} \alpha_1 + 3\alpha_2 + \beta &> 0, & 2\alpha_1 + \alpha_2 + \beta &> 0, & \alpha_1 + \alpha_2 + \beta &> 0, \\ 5\alpha_1 + \alpha_2 + \beta &\leq 0, & 5\alpha_1 + 2\alpha_2 + \beta &\leq 0, & 2\alpha_1 + 7\alpha_2 + \beta &\leq 0. \end{aligned}$$

Одно из решений: $\alpha_1 = -6, \alpha_2 = -4, \beta = 24$, т.е.
 $f(x) = -6\xi_1 - 4\xi_2 + 24$.

Предположим теперь, что имеется некоторый новый вид сырья, характеризующийся вектором $(3,3)$. Так как $f(3,3) = -6 < 0$, то мы делаем прогноз, что это сырье технологически недопустимо.

3.6. Геометрический смысл

Если $\{x: f(x)=0\}$ представляет собой поверхность в R^n , то задача дискриминантного анализа заключается в построении поверхности, разделяющей A' и B' ; эта поверхность делит всё пространство R^n на два образа.

В случае, когда f — аффинная функция, $f(x)=\langle c, x \rangle + \beta$, речь идет о построении разделяющей гиперплоскости с нормалью c .

3.7. Задача дискриминантного анализа для k образов

Пусть имеется неизвестное нам разделение пространства R^n на k образов:

$$R^n = A_1 \cup \dots \cup A_k, \quad A_i \cap A_j = \emptyset \quad (i \neq j).$$

Известны подмножества $A'_i \subset A_i$ ($i=1, \dots, k$).

Приведем один из способов сведения задачи дискриминантного анализа для k образов к последовательности задачи с двумя образами.

Построим $k-1$ разделяющих функций f_1, \dots, f_{k-1} , каждая из которых делит два множества: f_1 отделяет A'_1 от $A'_2 \cup \dots \cup A'_k$; f_2 отделяет A'_2 от $A'_3 \cup \dots \cup A'_k$; f_{k-1} отделяет A'_{k-1} от A'_k . Таким образом, для всякого $j \in \overline{1, k-1}$ решаем задачу: найти

$$f_j \in F: f_j(x) > 0 \quad (\forall x \in A'_j), \quad f_j(x) < 0 \quad (\forall x \in \bigcup_{i>j} A'_i).$$

Если функции f_1, \dots, f_{k-1} , удовлетворяющие указанным условиям, найдены, то классификация любого $y \in R^n$ производится согласно правилу

$$y \in \begin{cases} A_1, & \text{если } f_1(y) > 0; \text{ иначе} \\ A_2, & \text{если } f_2(y) > 0; \text{ иначе} \\ \dots\dots\dots \\ A_{k-1}, & \text{если } f_{k-1}(y) > 0; \text{ иначе} \\ A_k \end{cases}$$

Итак, мы имеем право в дальнейшем изучать методы только для задач дискриминантного анализа с двумя образами.

3.8. Пример

Пусть $x = (\xi_1, \xi_2)$ есть вектор состояния сырья некоторого производства, где ξ_i - процентное содержание i -го химического компонента в единице сырья ($i=1,2$). Требуется построить правило прогноза качества сырья на основании материала наблюдений: A'_1 - множество, содержащее некоторые векторы состояния хорошего сырья, A'_2 содержит некоторые векторы, отвечающие среднему качеству, A'_3 - аналогичное множество векторов для плохого качества сырья. Пусть, например,

$$A'_1 = \{x_1 = (2, 4); x_2 = (2, 2)\}, A'_2 = \{x_3 = (5, 2)\}, A'_3 = \{x_4 = (7, 1)\}.$$

Предположим, F - класс всех линейных функций. Решаем 2 задачи дискриминации:

1. $f_1 \in F: f_1(x) > 0 \ (\forall x \in A_1'), f_1(x) < 0 \ (\forall x \in A_1' \cup A_3')$;
2. $f_2 \in F: f_2(x) > 0 \ (\forall x \in A_2'), f_2(x) < 0 \ (\forall x \in A_3')$.

Одно из решений этих задач: $f_1(x) = -2\xi_1 + 3\xi_2$, $f_2(x) = -\xi_1 + 5\xi_2$.

Пусть для прогноза представлены три вектора состояния сырья: $y_1 = (1, 3)$, $y_2 = (12, 2)$, $y_3 = (4, 2)$.

Так как $f_1(y_1) > 0$, то делаем прогноз y_1 отвечает сырью хорошего качества.

Так как $f_1(y_2) < 0$, $f_2(y_2) < 0$ то y_2 согласно прогнозу есть вектор состояния сырья плохого качества.

Далее $f_1(y_3) < 0$, $f_2(y_3) > 0$, поэтому y_3 относим к сырью среднего качества.

3.9. Геометрический смысл

Легко видеть, что пример, описанный в предыдущем пункте, порождает следующую картину в R_+^2 :

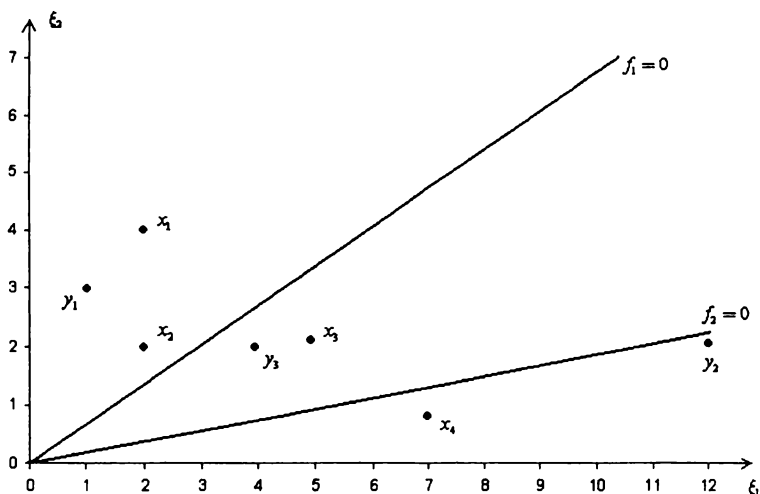


Рис. 2.

§ 4. Модель таксономии

4.1. Задача таксономии

Пусть задано конечное множество

$$X = \{x_1, \dots, x_r\} \subset R^n,$$

и пусть k - число классов (таксонов), на которые нужно разбить множество X :

$$X = X_1 \cup \dots \cup X_k, \quad X_i \cap X_j = \emptyset \quad (i \neq j); \quad (1)$$

здесь x_i - i -й таксон. Число k может быть либо задано заранее, либо должно быть определено в результате решения задачи таксономии.

Разбиение (1) может быть осуществлено многими способами. Если имеется функционал φ , оценивающий качество разбиения:

$$\varphi(X_1, \dots, X_k),$$

причем большему значению φ соответствует лучшее качество, то задача таксономии имеет вид:

$$\max \{ \varphi(X_1, \dots, X_k) : X = X_1 \cup \dots \cup X_k, \quad X_i \cap X_j = \emptyset \quad (i \neq j) \}.$$

Функционал φ может учитывать качество строения таксона $x_i(\forall_i)$, т.е. его однородность, достаточную близость друг к другу его элементов, и достаточную степень различия между таксонами с различными номерами, и какие-либо другие критерии.

Многие известные постановки задач таксономии диктуются алгоритмами, реализующими их решение. Приведем некоторые конкретные постановки.

1. Задан класс F функций $f(x)$. Требуется найти все максимально совместные подсистемы (μ - подсистемы) для системы

$$f(x) \leq 0 \quad (\forall x \in X), f \in F \quad (2)$$

Если система $f(x) \leq 0 \quad (\forall x \in X_1)$ есть μ - подсистема системы (2), то X_1 считается таксоном множества X .

2. Пусть $r(x, y)$ – неотрицательная функция двух векторов $x, y \in R^n$, характеризующая их «близость» друг к другу, причем $r(x, x) = 0$. Требуется чтобы

$$\forall i = 1, k, \quad \forall x, y \in X_i : r(x, y) \leq R; \quad \forall j, \quad \forall i \neq j, \quad \forall x \in X_i, \quad \forall y \in X_j : r(x, y) > R.$$

3. Разбиение X на два таксона: $X_1 \subset X$ такое, что разделяющий X_1 и $X \setminus X_1$ плоский слой наибольшей толщины является максимально (по всем возможным $X_1 \subset X$) толстым.

4.2. Примеры практических задач таксономии

В качестве примеров задачи таксономии можно привести задачи создания типологии объектов того или иного класса (социологических групп, предприятий, технических и природных систем и т.д.).

Приведем пример решения задачи в постановке 1 (см. рис. 3).

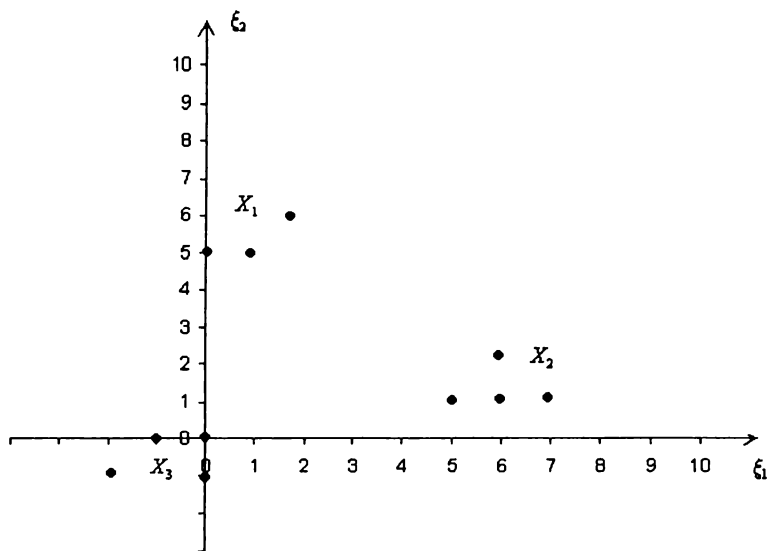


Рис. 3.

Пусть $X = \{x_1 = (0, 0), x_2 = (0, 5), x_3 = (1, 5), x_4 = (2, 6), x_5 = (-1, 0), x_6 = (-2, -1),$

$x_7 = (0, -1), x_8 = (5, 1), x_9 = (6, 1), x_{10} = (6, 2), x_{11} = (7, 1)\}$,

$F = \{f(x) = (\xi_1 - \alpha_1)^2 + (\xi_2 - \alpha_2)^2 - 5 : \alpha_1 \in R^1, \alpha_2 \in R^1\}$.

Таксоны, соответствующие μ -подсистемам системы (2), показаны на рис. 3.

§ 5. Выбор признакового пространства

5.1. Постановка задачи выбора наиболее информативной подсистемы признаков

Приведем одну из постановок. Пусть заданы $A, B \subset R^n$ и $F \subset \{R^n \rightarrow R^1\}$, т.е. все элементы задачи дискриминантного анализа, Предположим также, что заданы класс допустимых преобразований признакового пространства:

$$G \subset \{R^n \rightarrow R^1\} \cup \{R^n \rightarrow R^2\} \cup \dots \cup \{R^n \rightarrow R^n\}$$

и класс допустимых преобразований множества разделяющих функций:

$$\Phi \subset \{F \rightarrow H \subset \{R^n \rightarrow R^1\} \cup \{R^{n-1} \rightarrow R^1\} \cup \dots \cup \{R^1 \rightarrow R^1\}\}$$

Пусть качество дискриминации множеств A, B классом F оценивается некоторым функционалом $\psi(A, B, F)$. Тогда в общем виде задача выбора подсистемы признаков такова: найти $g \in G$ и $\varphi \in \Phi$ такие что, $\psi(A_g, B_g, F_\varphi)$ принимает наиболее возможное значение ; здесь g и φ согласованы так, что если $g \in \{R^n \rightarrow R^n\}$, то $F_\varphi \subset \{R^n \rightarrow R^1\}$. При этом ещё накладывается то или иное дополнительное условие на систему признаков (часто ограничение сверху на число используемых признаков).

Можно, например, так конкретизировать эту постановку. Пусть $I = \overline{1, n}$, $x = (\xi_1, \dots, \xi_n) = x(I)$, а если $J = \{i_1, \dots, i_r\} \subset I$, то $x(J) = (\xi_{i_1}, \dots, \xi_{i_r})$.

Пусть G - множество всех преобразований типа $x \rightarrow x(J)$ ($\forall J \subset I$), F - класс всех аффинных функций, $F\Phi$ - класс всех аффинных функций, $A(J) = \{x(J) : x \in A\}$.

Одна из простейших задач выбора признаков имеет вид:

$\min \{|J| : \text{множества } A(J) \text{ и } B(J) \text{ аффинно разделимы}\}.$

Ещё более простая постановка: $\min \{|J| : A(J) \cap B(J) = \emptyset\}.$

Если $\omega(J)$ - качество решения задачи дискриминации при замене A и B на $A(J)$ и $B(J)$, то можно рассматривать задачи:

$\max \{\omega(J) : |J| = q\}$ или $\min \{|J| : \omega(J) \geq \alpha\}.$

5.2. Примеры практических задач

Как сказано выше, математической моделью объекта в распознавании образов является n -мерный вектор $x = (\xi_1, \dots, \xi_n)$. Параметры, значениями которых являются числа ξ_1, \dots, ξ_n , называются признаками объекта. Для характеристики объекта можно выбирать различные системы признаков, и задача оценки информативности как раз состоит в нахождении оптимальной (с точки зрения того или иного критерия оптимальности) системы признаков.

Например, может быть поставлена задача: найти систему признаков, измерение которых связано с наименьшими возможными затратами, но при использовании которых возможно разделение заданных конечных множеств объектов с помощью функций заданного класса. Задача с другим критерием оптимальности: выбрать подсистему, включающую m признаков из общего их числа n , $n > m$, обеспечивающую наилучшее качество разделения.

Приведем конкретный числовой пример. Пусть заданы два множества

A и B ,

дискриминацию которых нужно произвести:

$$A = \{(1, 2), (7, 4), (0, -3)\}, B = \{(6, 3), (1, 5)\}.$$

Требуется выбрать подсистему из одного признака: $J \subset \overline{1,2}, |J|=1$. Если выбрать подсистему $J_1 = \{1\}$, то вместо множеств A и B мы должны разделить множества $A(J_1) = \{(1), (7), (0)\}$, $B(J_1) = \{(6), (1)\}$, что невозможно ввиду $A(J_1) \cap B(J_1) = \{(1)\} \neq \emptyset$.

Если же выбрать $J_2 = \{2\}$, то $A(J_2) = \{(2), (4), (-3)\}$, $B(J_2) = \{(3), (5)\}$, и ввиду $A(J_2) \cap B(J_2) = \emptyset$ разделение множеств $A(J_2)$ и $B(J_2)$ возможно, например, функцией $f(\xi)$, график которой приведен на рис. 4.

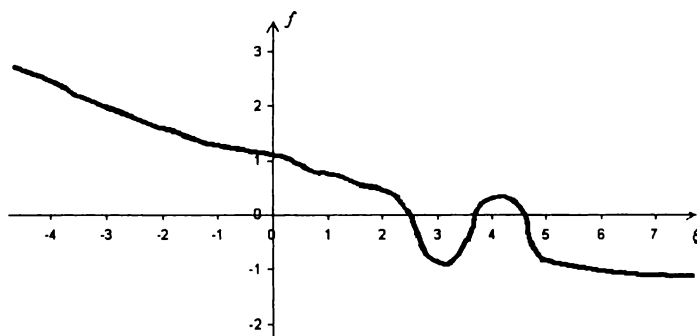


Рис. 4.

Содержательными примерами задач оценки информативности признаков могут служить задачи выбора значащих факторов при изучении тех или иных эмпирических зависимостей, в том числе задача выбора зонообразующих факторов в географии.

Глава 2.

ЛИНЕЙНЫЙ ДИСКРИМИНАНТНЫЙ АНАЛИЗ

§ 1. Методы дискриминантного анализа

Напомним формулировку задачи дискриминантного анализа.

Заданы: $A, B \subset R^n$; класс функций: $F \subset \{R^n \rightarrow R\}$; условие нормализации $\varphi(A, B, f), f \in F$.

Требуется найти функцию f n -мерного вектора x , удовлетворяющего условиям: $f \in F; f(x) > 0 (\forall x \in A), f(x) < 0 (\forall x \in B)$ (или может быть условие $f(x) \leq 0 (\forall x \in B)$); $\varphi(A, B, f) \rightarrow \max$.

Рассмотрим пока более простую формулировку:

$$\left. \begin{array}{l} f \in F; \\ f(x) > 0 (\forall x \in A); \\ f(x) < 0 (\forall x \in B). \end{array} \right\} \quad (1)$$

Зададимся следующими вопросами:

1. При каких условиях задача (1) сводится к решению конечной системы линейных неравенств?
2. Каковы принципы выбора класса F ?

1.1 Сведение к линейным неравенствам

Ответ на первый вопрос достаточно прост. Чтобы система (1) была конечной, казалось бы, достаточно, чтобы A и B были конечны, но на самом деле конечности множеств A и B вообще говоря, недостаточно для конечности системы (1), так как есть ещё условие $f \in F$, а задание класса F может быть и довольно сложным.

Система (1) линейна в том и только том случае, когда существует конечный набор функций $f_0(x), f_1(x), \dots, f_m(x)$ такой, что всякая функция $f \in F$ может быть представлена в виде:

$$f(x) \equiv \sum_{i=1}^m y_i f_i(x) + f_0(x). \quad (2)$$

1.2. Принципы выбора класса разделяющих функций

Перейдем к вопросу 2. Этот вопрос можно следующим образом связать с вопросом 1. Выше уже указано, что мы руководствуемся эвристическим принципом: чем проще структура функции разделяющая множества A и B , тем лучше. Поэтому если бы «простота» измерялась некоторым функционалом φ , так что $\varphi(f) > \varphi(g)$ означало бы, что функция f проще функции g , можно было бы выбрать достаточно широкий класс F (например, класс всех полиномов) и ставить задачу в виде:

$$\max \{ \varphi(f) : f \in F, f(x) > (\forall x \in A), f(x) < (\forall x \in B) \}.$$

Однако это слишком сложная задача, если к тому же учесть обстоятельство, что в реальных задачах распознавания образов материал обучения представляет собой последовательность, число членов которой растёт в дискретном времени, и поэтому метод решения задачи должен быть итерационным.

Поэтому рассмотрим другой подход. Будем рассматривать только функции вида (2), что позволит нам иметь дело только с системами линейных неравенств. Тогда можно считать, что простота функций (2) = простота наиболее сложной из f_i . Класс F задается так:

$$F = \left\{ f(x) = \sum_{i=1}^m y_i f_i(x) + f_0(x) : y_i \in R, f_i \in \Phi (\forall i \in \overline{1, m}) \right\}.$$

При этом само число m может быть либо фиксированным ($F = F(m, \Phi), m \in \{1, 2, \dots\}$), либо произвольным: $F = F(\Phi)$.

Будем рассматривать классы возрастающей сложности (при $x = (\xi_1, \dots, \xi_n)$):

$$\begin{aligned} \Phi_1 &= \{\xi_1, \dots, \xi_n\}; \\ \Phi_2 &= \Phi_1 \cup \{\xi_i \xi_j : i, j = 1, \dots, n\}; \\ \Phi_3 &= \Phi_2 \cup \{\xi_i \xi_j \xi_k : i, j, k = 1, \dots, n\}; \\ &\dots\dots\dots \\ \Phi_p &= \Phi_{p-1} \cup \{\xi_{i_1} \xi_{i_2} \dots \xi_{i_p} : i_1, \dots, i_p = 1, \dots, n\} \\ &\dots\dots\dots \end{aligned}$$

Классу Φ_1 соответствует класс F аффинных функций, Φ_2 класс F квадратичных функций, ..., Φ_p - класс F полиномов p -й степени.

1.3. Требования к алгоритмам дискриминантного анализа

Проблема состоит в следующем. Желательно выбрать i такое, что при классе Φ_i (т.е. при $F = F(\Phi_i)$) система (1) разрешима, а при классе Φ_{i-1} она неразрешима.

Примой подход (перебор классов Φ_1, Φ_2, \dots один за другим) слишком неэкономичен; было бы желательно хотя бы уметь использовать результаты расчетов для класса Φ_{i-1} при переходе к классу Φ_i .

Несколько другой подход, позволяющий естественное последовательное наращивание сложности разделяющего

класса, представляет метод комитетов, о котором будет сказано позднее.

Хорошо было бы придумать процедуру, в которой построение разделяющей функции сопровождалось добавлением (по мере необходимости) новых членов вида $\nu \prod_{i=1}^p \xi_i$. Например, было бы естественным применять процедуру, приводящую к чебышевскому приближению, и это приближение брать как стартовое при решении аналогичной задачи на следующем шаге расширения класса Φ . Этот подход хорош и в случае расширяющегося материала обучения.

Ещё один подход: использовать следующий итерационный метод для получения кусочно-линейной разделяющей поверхности (см. рис. 5).

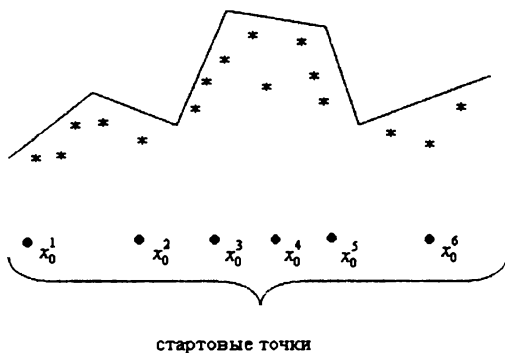


Рис. 5.

Каждая из стартовых точек дает начало итерационной последовательности (получаемой методом линейной коррекции или методом чебышевских приближений), сходящейся к своему куску поверхности. «Свой» кусок обучающего множества для каждой последовательности $\{x_k\}_{k=0}^{+\infty}$ определяется близостью к этой последовательности. Так, для каждой последующей точки обучающего множества $A \cup B$,

поступающей на вход распознающей системы, определяется ближайшая к ней из множества $\{x_i^1, \dots, x_i^p\}$, и относительно этой точки делаем коррекцию.

Эта конструкция относится к классу коллективных или размытых решений (когда вместо одной точки – множество), как в методе комитетов. Обоснования этой процедуры пока нет. Теперь можно перейти к изложению конкретных методов решения систем линейных неравенств, понимая в свете сказанного их роль в дискриминантном анализе, т.е. анализируя их с точки зрения приспособляемости к возможной несовместности и к возможному расширению материала обучения.

§ 2. Разделяющие возможности аффинных функций

2.1. Число аффинных дихотомий

Всякому конечному множеству объектов можно ставить в соответствие различные разбиения его на классы – в зависимости от цели, с которой проводится классификация объектов. Так, если на множество всех людей смотреть с точки зрения того, какими болезнями они болеют, то они получают одну классификацию (медицинская диагностика), а с социальной точки зрения они будут делиться на другого рода классы (социальная стратификация).

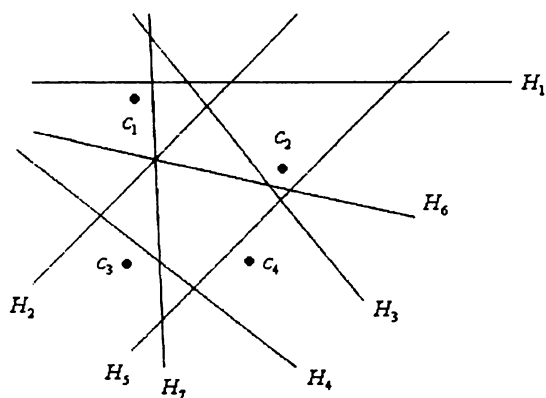
Поэтому каждый класс F разделяющих функций, который мы употребляем для решения задач РО, естественно оценивать по такому показателю: сколько разных дихотомий одного и того же конечного множества может быть осуществлено с помощью дискриминантных функций этого класса.

Дихотомия множества M – разбиение множества M на два класса: $M = M_1 \cup M_2, M_1 \cap M_2 = \emptyset$.

Введем обозначение $L(m, n)$ – число дихотомий множества состоящего из m элементов пространства R^n , осуществимых с помощью аффинных дискриминантных функционалов (АДФ).

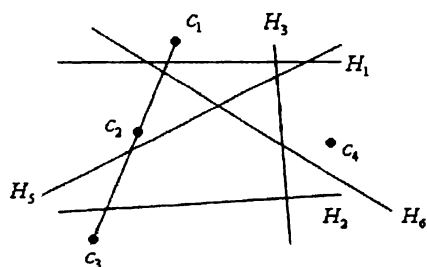
Для каждого разбиения множества M на два подмножества существует две различные классификации: если $M = M_1 \cup M_2, M_1 \cap M_2 = \emptyset$, то можно считать M_1 - первым классом, а M_2 - вторым, либо M_2 - первым классом, а M_1 - вторым. Поэтому $L(m, n) = 2$ [число разбиений множества M на два подмножества одной гиперплоскостью].

Приведем пример для $m=4, n=2$ (см. рис. 6). Число всех дихотомий равно $C_m^0 + C_m^1 + C_m^2 + \dots + C_m^m = 2^m = 16$.



Класс	№	Аффинные
1	1	*
2	2	*
3	3	*
4	4	*
1,2	5	*
1,3	6	
1,4	7	
2,3	8	
2,4	9	*
3,4	10	*
1,2,3	11	*
1,2,4	12	*
1,3,4	13	*
2,3,4	14	*
1,2,3,4	15	*
\emptyset	16	*

Рис.6.



Класс	№	Аффинные
1	1	*
2	2	
3	3	*
4	4	*
1,2	5	*
1,3	6	
1,4	7	*
2,3	8	*
2,4	9	
3,4	10	*
1,2,3	11	*
1,2,4	12	*
1,3,4	13	
2,3,4	14	*
1,2,3,4	15	*
\emptyset	16	*

Рис. 7.

Взаимное расположение точек в R^2 не влияет на величину $L(4,2)=14$, если только такие три точки не являются лежащими на одной прямой.

Если, например, точки c_1, c_2, c_3 лежат на одной прямой (как на рис. 7), причем c_2 лежит между c_1 и c_3 , то нельзя отделить точку c_2 от всех остальных. Число всех возможных дихотомий равно 12.

2.2. Расположение точек общего вида

Определение. Расположение m точек в R^n называется расположением общего вида, если $\forall k, k=1, 2, \dots, n$: всякое подмножество, состоящее из $(k+1)$ -й точки, не лежит в $(k-1)$ -мерном линейном многообразии.

Теорема. Если некоторые m точек в R^n имеют расположение общего вида, то $L(m, n) = L(m-1, n) + L(m-1, n-1)$ (1)

Доказательство. Пусть $R^n \supset M = \{c_1, \dots, c_m\}$ - множество точек, имеющее расположение общего вида, $M' = \{c_1, \dots, c_{m-1}\}$; $L_{c_m}(m-1, n)$ - число аффинных классификаций множества M' , получаемых с помощью гиперплоскостей, содержащих c_m . Тогда число всех аффинных классификаций M' , которые нельзя получить с помощью таких плоскостей, равно $L(m-1, n) - L_{c_m}(m-1, n)$.

Каждая из классификаций последнего вида дает только одну классификацию множества M . А каждой классификации множества M' , полученной с помощью плоскости проходящей через c_m , соответствуют две классификации множества M : при одной c_m относится к одному классу, а при другой - ко второму (это делается через небольшое перемещение плоскости). Тогда

$$L(m, n) = L(m-1, n) - L_{c_m}(m-1, n) + 2L_{c_m}(m-1, n) = L(m-1, n) + L_{c_m}(m-1, n).$$

Покажем, что

$$L_{c_m}(m-1, n) = L_{c_m}(m-1, n-1). \quad (2)$$

Для этого построим отрезки $[c_m, c_i] (i=1, \dots, m-1)$. Каждые два отрезка пересекаются только в точке c_m , так как каждые три точки имеют расположение общего вида. Пусть H — гиперплоскость, пересекающая все прямые, каждая из которых содержит какой-нибудь отрезок.

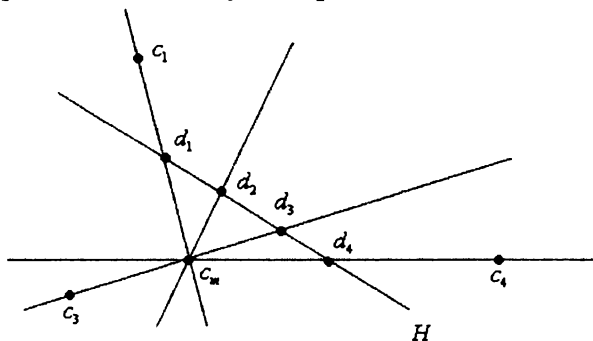


Рис.8.

Обозначим через $N = \{d_1, \dots, d_{m-1}\}$ множество точек пересечения гиперплоскости H со своими прямыми. Число классификаций множества M образуемых $(n-1)$ - мерными плоскостями, проходящими через точку c_m , равно числу классификаций множества N , образуемых $(n-2)$ - мерными гиперплоскостями в H . Но H есть $(n-1)$ - мерное многообразие. При этом, поскольку никакая совокупность $k+1$ точек, включающая c_m , не лежит в $(k-1)$ -мерном многообразии, то в H расположение точек общего вида. Поэтому соотношение (2) доказано, а вместе с ним и теорема.

Замечание. Формула (1) дает возможность составить таблицу значений $L(m, n)$,

Таблица значений $L(m, n)$

m/n	1	2	3	4	5
1	2	2	2	2	2
2	4	4	4	4	4
3	6	8	8	8	8
4	8	14	16	16	16
5	10	22	30	32	32
6	12	32	52	62	64
7	14	44	84	114	126
8	16	58	128	198	240

Достаточно вписать первую строку (она состоит из чисел 2), а также столбец (2 m), а затем применять формулу (1).

§ 3. Метод линейной коррекции

Распознающая система имеет вид:

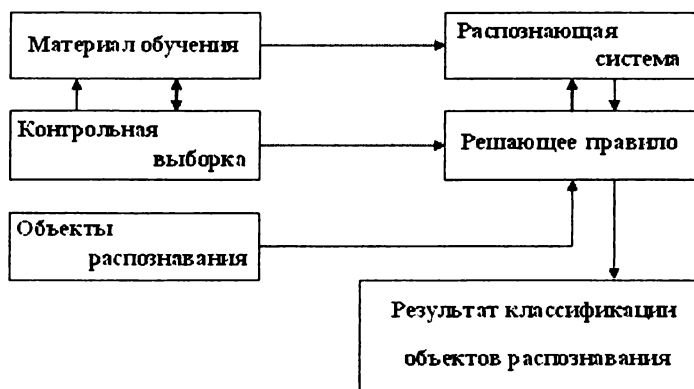


Рис. 9.

3.1. Алгоритм линейной коррекции

Требуется найти $x \in R^n, y \in R^1$ $\langle x, c \rangle - y > 0 (\forall c \in A'), \langle x, c \rangle - y < 0 (\forall c \in B')$.

Введем

обозначения:

$$(c; -1) = a; (x; y) = z; \{(c; -1) : c \in A'\} = A; \{(c; -1) : c \in B'\} = B.$$

Тогда надо решить систему

$$\langle a, z \rangle > 0 (\forall a \in A), \langle a, z \rangle < 0 (\forall a \in B) \quad (3)$$

Составим из элементов $A \cup B$ бесконечную последовательность такую, что в ней каждое $a \in A \cup B$ встречается бесконечное число раз. Обозначим эту последовательность через $\{a^1, a^2, \dots\}$.

Выбираем z^1 произвольно. Строим последовательность $\{z^k\}_{k=1}^{+\infty}$ согласно правилу

$$z^{k+1} = \begin{cases} z^k, & \text{если } \langle z^k, a^k \rangle > 0, a^k \in A; \\ z^k, & \text{если } \langle z^k, a^k \rangle > 0, a^k \in B; \\ z^k - \alpha a^k, & \text{если } \langle z^k, a^k \rangle \geq 0, a^k \in B; \\ z^k + \alpha a^k, & \text{если } \langle z^k, a^k \rangle \leq 0, a^k \in A, \end{cases}$$

где $\alpha > 0$.

Этот алгоритм называется алгоритмом линейной коррекции.

Теорема Новикова.

Теорема. Пусть множества A' и B' аффинно разделимы. В результате применения алгоритма линейной коррекции возникает последовательность $\{z^k\}_{k=1}^{+\infty}$ такая, что $\exists T: z^T = z^{T+1} = \dots, z^T$ - решение системы (3).

Доказательство. Не ограничивая общности, можно полагать, что $\alpha=1$ и что мы решаем систему $\langle a, z \rangle > 0 (\forall a \in A)$ с помощью следующего метода.

Из множества A составлена бесконечная последовательность $\{z^k\}_{k=1}^{+\infty}$, где каждый элемент множества A встречается бесконечное число раз. Последовательность $\{z^k\}_{k=1}^{+\infty}$ строим согласно соотношению

$$z^{k+1} = \begin{cases} z^k, & \text{если } \langle z^k, a^k \rangle > 0; \\ z^k + a^k, & \text{если } \langle z^k, a^k \rangle \leq 0. \end{cases}$$

Кроме того, исключим из последовательности $\{a^k\}_{k=1}^{+\infty}$ те объекты, для которых $z^{k+1} = z^k$, т.е. $\langle z^k, a^k \rangle > 0$. Пусть $\{a^k\}_{k=1}$ - последовательность после такого исключения. Возьмем в качестве z^0 вектор $\theta = (0, \dots, 0)$. Тогда

$$z^k = a^1 + \dots + a^k \quad (4)$$

Фиксируем некоторое \bar{z} такое, что $\langle \bar{z}, a \rangle > 0 \ (\forall a \in A)$. Положим $\varepsilon = \min \{ \langle \bar{z}, a \rangle : a \in A \} > 0$. Умножим скалярно левую и правую части соотношения (4) на вектор \bar{z} :

$$\langle z^k, \bar{z} \rangle = \langle a^1, \bar{z} \rangle + \dots + \langle a^k, \bar{z} \rangle \geq k\varepsilon.$$

Так как по неравенству Коши-Шварца, $\|z^k\|^2 \geq \langle z^k, \bar{z} \rangle^2 / \|\bar{z}\|^2$, то

$$\|z^k\|^2 \geq k^2 \varepsilon^2 \|\bar{z}\|^{-2} \quad (5)$$

С другой стороны, так как $z^{j+1} = z^j + a^j$, то $\|z^{j+1}\|^2 = \|z^j\|^2 + 2\langle z^j, a^j \rangle + \|a^j\|^2$. Но $\langle z^j, a^j \rangle \leq 0$, следовательно $\|z^{j+1}\|^2 - \|z^j\|^2 \leq \|a^j\|^2 \ (\forall j)$.

Сложим последние неравенства полученные для $j = 1, \dots, k$:

$$\|z^{k+1}\|^2 \leq \sum_{j=1}^k \|a^j\|^2 + \|z^1\|^2.$$

Полагая $M = \max \{ \|a^j\|^2 : a \in A \}$, получим:

$$\|z^{k+1}\|^2 \leq k \cdot M + \|z^1\|^2. \quad (6)$$

Неравенство (6) противоречит неравенству (5) при больших k , следовательно, последовательность $\{a^k\}$ обрывается через конечное число номеров.

Теорема доказана.

§ 4. Методы линейного программирования

4.1. Симплекс-метод

Пусть в задаче дискриминантного анализа требуется найти линейную разделяющую функцию $f(x) = z_1 x_1 + \dots + z_n x_n$ для множеств $A, B \in R^n$. Здесь z_i – искомые коэффициенты разделяющей функции. Пусть

$$A = \{a_j = (a_{j1}, \dots, a_{jn}) : j \in \overline{1, s}\}, B = \{a_j = (b_{j1}, \dots, b_{jn}) : j \in \overline{1, t}\}.$$

Задача дискриминантного анализа принимает вид:

$$\left. \begin{aligned} a_{j1} z_1 + \dots + a_{jn} z_n &> 0 \quad (j \in \overline{1, s}), \\ b_{j1} z_1 + \dots + b_{jn} z_n &> 0 \quad (j \in \overline{1, t}). \end{aligned} \right\} \quad (1)$$

Рассмотрим также систему

$$\left. \begin{aligned} a_{j1} z_1 + \dots + a_{jn} z_n &\geq 1 \quad (j \in \overline{1, s}), \\ b_{j1} z_1 + \dots + b_{jn} z_n &\leq -1 \quad (j \in \overline{1, t}). \end{aligned} \right\} \quad (2)$$

Очевидно, что если система (1) совместна, то совместна и система (2); при этом всякое решение системы (2) является решением системы (1).

В то же самое время решения системы (2), к которой можно добавить линейный критерий оптимальности, имеются стандартные методы линейного программирования, в том числе симплекс метод которым и можно воспользоваться.

4.2. Методы фейеровского типа

Рассмотрим итерационные методы решения системы линейных неравенств

$$L_j(x) = \langle c_j, x \rangle - \alpha_j \leq 0 \quad (\forall j \in \overline{1, m}), \quad (1)$$

основанные на использовании фейеровских отображений. Будем предполагать, что $\|c_j\| \neq 0 \quad (\forall j \in \overline{1, m})$.

Фейеровские методы для решения систем линейных и выпуклых неравенств разработаны и исследованы в работах И.И. Ерёмкина. Отметим, что если имеется метод получения как угодно точного решения системы (1), т.е. $\forall \varepsilon > 0, \exists x(\varepsilon) = \bar{x}$, где $x(\varepsilon)$ - элемент, получаемый в результате применения метода, причем $\max_{j \in \overline{1, m}} L_j(\bar{x}) \leq \varepsilon$, то на основе этого метода можно построить метод получения точного решения систем строгих неравенств

$$L_j(x) < 0 \quad (\forall j \in \overline{1, m}), \quad (2)$$

если последняя совместна. Способ построения такого метода, предложенный И.И. Ереминым, состоит в следующем.

Взяв $\delta > 0$, запишем следующую систему относительно $z = (x; y) \in R^{n+1}$:

$$\begin{aligned} L_j(z) &= \langle c_j, x \rangle - \alpha_j y + \delta \leq 0 \quad (\forall j \in \overline{1, m}), \\ L_0(z) &= -y + \delta \leq 0. \end{aligned} \quad (3)$$

Замечание. Система (2) совместна тогда и только тогда, когда совместна система (3), причем если $\bar{z} = (\bar{x}; \bar{y})$, удовлетворяет неравенствам $L_j(z) \leq \varepsilon \quad (j = 0, 1, \dots, m)$, где $0 < \varepsilon < \delta$, то \bar{x}/\bar{y} - решение системы (2).

Следовательно, для нахождения решения системы (2) достаточно иметь метод получения приближенных решений системы линейных неравенств вида (1).

Доказательство. Пусть система (2) совместна и \bar{x} - некоторое её решение. Положим $x' = y'\bar{x}$, $y' = \left\{ \max_{j \in \overline{1, m}} \frac{\delta}{\alpha_j - \langle c_j, \bar{x} \rangle}, \delta \right\}$.

Тогда $y' \geq \delta'$, причем $\langle c_j, x' \rangle - \alpha_j y' + \delta = y' [\langle c_j, \bar{x} \rangle - \alpha_j] + \delta \leq \frac{\delta}{\alpha_j - \langle c_j, \bar{x} \rangle} [\langle c_j, \bar{x} \rangle - \alpha_j] + \delta \leq 0 \ (\forall j \in \overline{1, m})$, что и требовалось.

Обратно, пусть $(x'; y')$ - некоторое решение системы (3):

$$\left. \begin{aligned} \langle c_j, x' \rangle - \alpha_j y' + \delta &\leq 0, \\ y' &\geq \delta. \end{aligned} \right\}$$

Тогда

$$\left. \begin{aligned} \left\langle c_j, \frac{x'}{y'} \right\rangle - \alpha_j + \frac{\delta}{y'} &\leq 0, \\ y' &\geq \delta. \end{aligned} \right\}$$

Следовательно, вектор $\bar{x} = x'/y'$ есть решение системы (2).

Пусть теперь $0 < \varepsilon < \delta$, причем для $\bar{z} = (\bar{x}; \bar{y})$: $L_j(\bar{z}) \leq \varepsilon \ (j = 0, 1, \dots, m)$.

Тогда

$$\begin{aligned} \langle c_j, \bar{x} \rangle - \alpha_j \bar{y} + \delta &\leq \varepsilon < \delta \ (j = 0, 1, \dots, m); \\ \bar{y} &\geq \delta > \varepsilon; \end{aligned}$$

отсюда $\left\langle c_j, \frac{\bar{x}}{\bar{y}} \right\rangle - \alpha_j \leq \frac{\varepsilon - \delta}{\bar{y}} \ (\forall j \in \overline{1, m})$, т.е. x'/y' - решение системы (2).

Пусть $\emptyset \neq M \subset R^n$. Отображение $\varphi \in \{R^n \rightarrow R^n\}$ называется *М-фейеровским*, если $\forall x \notin M, \forall y \in M : [\varphi(y) = y] \& \|\varphi(x) - y\| < \|x - y\|$.

Если множество M допускает хотя бы одно M -фейеровское отображение, то оно выпукло и замкнуто. Если φ непрерывно, то $\forall x \in R^n, \exists \bar{x} \in M : \varphi^k(x) \rightarrow \bar{x} (k \rightarrow +\infty)$.

Положим далее $M = \{x : (1)\}$. Имеется множество способов построения M -фейеровских отображений. Приведем некоторые из них.

Функции $d \in \{R^n \rightarrow R^1\}$ и $e \in \{R^n \rightarrow R^n\}$ образуют разделяющую пару для M если

- 1) $\forall x \notin M, \forall y \in M : \langle e(x), y - x \rangle + d(x) \leq 0$;
- 2) $\{y : d(y) \leq 0\} = M$.

Легко доказать, что если $(d; e)$ - разделяющая пара для M , то отображение

$$\varphi(x) = x - \alpha \frac{d^+(x)}{\|e(x)\|^2} e(x),$$

где $\alpha \in (0, 2)$ и $d^+(x) = \max\{0, d(x)\}$, является M -фейеровским. Некоторые примеры M -разделяющих пар:

- 1) $d(x) = \max_j L_j(x) = L_{j_*}(x); e(x) = c_{j_*}$;
- 2) $d(x) = \sum_{j=1}^m L_j^+(x); e(x) = \sum_{j \in S(x)} c_{j_*}$, где $S(x) = \{j : L_j(x) > 0\}$ и если $S(X) = \emptyset$, то $e(x) = 0$;
- 3) $d(x) = \sum_{j=1}^m L_j^{+2}(x); e(x) = \sum_{j=1}^m L_j^+(x) c_{j_*}$.

4.3. Метод нахождения чебышевского приближения

Если множества A и B неразделимы гиперплоскостью, то можно говорить о приближенном решении соответствующей задачи аффинного дискриминантного

анализа, конкретно — о нахождении чебышевского приближения возникающей в этой задаче систем линейных неравенств.

На этом пути мы можем прийти к следующей постановке.

Рассмотрим систему неравенств

$$\begin{aligned} -\langle a, x \rangle + 1 &\leq 0 \quad (\forall a \in A), \\ \langle b, x \rangle - 1 &\leq 0 \quad (\forall b \in B). \end{aligned}$$

Предположим, что эта система несовместна, определим её чебышевское уклонение:

$E = \min \{t \geq 0 : -\langle a, x \rangle + 1 \leq t \quad (\forall a \in A); \langle b, x \rangle - 1 \leq t \quad (\forall b \in B)\}$, а затем множество $M(E)$ чебышевских приближений:

$$M(E) = \{x : -\langle a, x \rangle + 1 \leq E \quad (\forall a \in A); \langle b, x \rangle - 1 \leq E \quad (\forall b \in B)\}.$$

Если имеется линейный критерий $\langle c, x \rangle$ качества разделяющей функции (подлежащей минимизации), то мы приходим к задаче $\min \{\langle c, x \rangle : x \in M(E)\}$.

Приведем один из алгоритмов решения этой задачи, не использующей предварительного определения величины E . Положим

$$\begin{aligned} \alpha(x) &= \max \{-\langle a, x \rangle + 1 : a \in A\}; \\ \beta(x) &= \max \{\langle b, x \rangle - 1 : b \in B\}; \\ \gamma(x) &= \max \{0, \alpha(x), \beta(x)\}; \\ \bar{\alpha}(x) &= -\arg \max \{-\langle a, x \rangle + 1 : a \in A\}; \\ \bar{\beta}(x) &= \arg \max \{\langle b, x \rangle - 1 : b \in B\}. \end{aligned}$$

Построим последовательность $\{x^k\}_{k=0}^{+\infty} \subset R^n$: x_0 — произвольная точка;

$$x_{k+1} = \begin{cases} x_k - \lambda_k \gamma(x_k) \bar{a}(x_k) \|\bar{a}(x_k)\|^{-2} + \varepsilon_k c, & \text{если } \gamma(x_k) = \lambda_k(x_k); \\ x_k - \lambda_k \gamma(x_k) \bar{b}(x_k) \|\bar{b}(x_k)\|^{-2} + \varepsilon_k c, & \text{если } \gamma(x_k) = \beta_k(x_k); \\ x_k - \varepsilon_k c, & \text{если } \gamma(x_k) = 0, \end{cases}$$

где $\varepsilon_k \rightarrow +0$, $\sum_{k=1}^{\infty} \varepsilon_k = +\infty$, $\lambda_k = \varepsilon_k > 0$ ($\forall k$).

Теорема. Пусть $E > 0$. Если $\tilde{M}(E) = \arg \min \{ \langle c, x \rangle : x \in M(E) \} \neq \emptyset$, $\|a\| = 1$ ($\forall a \in A$), $\|b\| = 1$ ($\forall b \in B$), $\varepsilon_k = \lambda_k \rightarrow +0$, $\sum \varepsilon_k = \infty$, множества A и B конечны, число $\|c\|$ достаточно мало, то $\lim_{k \rightarrow +\infty} \|x_k - \tilde{M}(E)\| = 0$.

Глава 3.

МЕТОД КОМИТЕТОВ В ДИСКРИМИНАНТНОМ АНАЛИЗЕ

§ 1. Комитетные конструкции

Как отмечено выше, задача нахождения дискриминантной функции f такой, что

$$f \in F, f(x) > 0 \ (\forall x \in A), f(x) \leq 0 \ (\forall x \in B), \quad (1)$$

может не обладать решениями даже в случае $A \cap B = \emptyset$, если класс F выбран недостаточно широким. Мы сталкиваемся, следовательно, с проблемой выбора класса F .

Однако можно предложить другой метод, который снимает эту проблему. Это метод комитетов, использующий вместо понятия решения системы (1) некоторое его обобщение, реализующее понятие «размытого» решения, т.е. множества $K \subset F$, обладающего теми или иными свойствами решения. Перейдем к конкретным определениям.

1.1. Комитет системы множеств

Определение. Пусть \mathcal{D} - некоторое семейство множеств D_j (произвольной природы): $\mathcal{D} = \{D_j : j \in J\}$, X - некоторое множество, и пусть $p \in [0, 1]$. p -комитетом в классе X для семейства \mathcal{D} называется конечное множество $K \subset X$ такое, что

$$|K \cap D_j| > p \cdot |K| \ (\forall j \in J),$$

где $|K|$ означает число элементов (мощность) множества K .

Комитетом в классе X для семейства \mathcal{D} называется множество $K \subset X$ (не обязательно конечное) такое, что $|K \cap D_j| > |K/D_j| (\forall j \in J)$.

Замечание. Если $X \cap \left(\bigcap_{j \in J} D_j\right) \neq \emptyset$, то любое множество $K \subset X \cap \left(\bigcap_{j \in J} D_j\right)$ является p -комитетом и комитетом.

1.2. Примеры комитетов систем множеств

Пример №1. $X = R^2$, $D_1 = \{x = (\xi_1, \xi_2) : \xi_1 > 0\}$, $D_2 = \{x = (\xi_1, \xi_2) : \xi_2 > 0\}$, $D_3 = \{x = (\xi_1, \xi_2) : \xi_1 + \xi_2 < 0\}$. Множество $K = \{x_1 = (1, 1), x_2 = (-2, 1), x_3 = (1, -2)\}$ является $2/3$ -комитетом в классе X для семейства $\mathcal{D} = \{D_1, D_2, D_3\}$. Множество K является также комитетом этого семейства.

Пример № 2. Пусть X есть четырехугольник $a_1 a_2 a_3 a_4$ в R^3 (см. рис. 10).

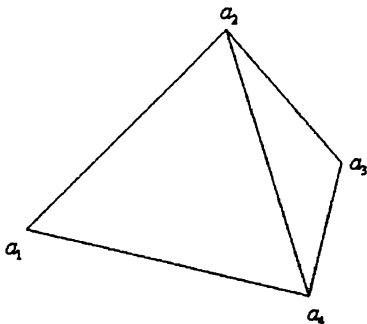


Рис. 10

Обозначим через D_1, \dots, D_4 грани четырехугольника X , через K - множество $\{a_1, a_2, a_3, a_4\}$. Легко убедиться в том, что K

есть 3/4-комитет X для семейства $\mathcal{D} = \{D_1, D_2, D_3, D_4\}$.
Одновременно k является и комитетом.

Пример №3. Пусть X – топологическое пространство, $a \in X$, \mathcal{D} – система окрестностей точки a , $K = \{a'\}_{i=1}^{+\infty}$ – последовательность, сходящаяся к элементу a .

Тогда k – комитет в классе X для семейства \mathcal{D} .

1.3. Определения и примеры комитетных конструкций

Определение. p -комитетом системы неравенств

$$\left. \begin{aligned} f_j(x) &> 0 (\forall j \in I_1), \\ f_j(x) &\geq 0 (\forall j \in I_2), \\ f_j(x) &= 0 (\forall j \in I_3) \end{aligned} \right\} \quad (2)$$

в классе X называется конечное множество $K \subset X$, являющееся p -комитетом системы множеств

$$\{x: f_j(x) > 0\} (j \in I_1), \{x: f_j(x) \geq 0\} (j \in I_2), \{x: f_j(x) = 0\} (j \in I_3).$$

Заметим, что комитет K системы (3) обладает свойством: каждому её неравенству удовлетворяют более половины членов комитета.

Пример. Рассмотрим систему неравенств над пространством R^2 :

$$\left. \begin{aligned} \xi_1 &> 0; \\ \xi_2 &\geq 0, \\ (\xi_1 + 1)^2 + (\xi_2 + 1)^2 &\leq 1,5. \end{aligned} \right\} \quad (3)$$

Очевидно, что система (3) несовместна. Пусть $X = R^2$, $K = \{x_1 = (1; 1), x_2 = (0; -1), x_3 = (-1, 2; 0, 1)\}$. Тогда K – комитет системы (3) в классе X .

Определение. p -комитетом системы неравенств относительно неизвестной функции f :

$$\left. \begin{aligned} f(x) &> 0 (\forall x \in X_1), \\ f(x) &\geq 0 (\forall x \in X_2), \\ f(x) &= 0 (\forall x \in X_3) \end{aligned} \right\} \quad (4)$$

В классе функций $F \subset \{R^n \rightarrow R^1\}$, где X_1, X_2, X_3 - известные подмножества пространства R^n , называется такое $K \subset F$, что множество K есть p -комитет системы множеств $\{f: f(x) > 0\} (x \in X_1), \{f: f(x) \geq 0\} (x \in X_2), \{f: f(x) = 0\} (x \in X_3)$.

Система (4) называется системой сопряженного вида.

Пример. Пусть $R^n = R^1$,

$$F = \{f_1(\xi) = \xi - 1, g_1(\xi) = -\xi + 1, f_2(\xi) = \xi - 2, g_2(\xi) = -\xi + 2, f_3(\xi) = \xi - 3, g_3(\xi) = -\xi + 3, \dots\}.$$

Рассмотрим систему сопряженного вида:

$$f(2) > 0, f(4) < 0, f(6) > 0, f(8) < 0, f(10) > 0.$$

Комитетом этой системы в классе F является следующее множество:

$$K = \{f_1, g_3, f_5, g_7, f_9\} \subset F.$$

Определение. Разделяющим заданные множества $A, B \subset R^n$ p -комитетом функций класса X называется p -комитет для системы неравенств $f(x) > 0 (\forall x \in A), f(x) \leq 0 (\forall x \in B)$.

Строго разделяющим называется p -комитет для системы $f(x) > 0 (\forall x \in A), f(x) < 0 (\forall x \in B)$.

Пример. Рассмотрим, как делит пространство R^2 на два образа строго разделяющий множества $A = \{(1, 1)\}$ и $B = \{(2, 0); (0, 2)\}$ комитет класса линейных функций (см. рис. 11).

Строим систему $f(1, 1) > 0, f(2, 0) < 0, f(0, 2) < 0$, где $f_1(\xi_1, \xi_2) = y_1 \xi_1 + y_2 \xi_2$, т.е. $y_1 + y_2 > 0, 2y_1 < 0, 2y_2 < 0$.

Комитет этой системы: $K = \{y^1 = (-1; -1), y^2 = (-1; 2), y^3 = (2; -1)\}$, т.е. разделяющий A и B комитет имеет вид: $\{f_1(x) = -\xi_1 - \xi_2, f_2(x) = -\xi_1 + 2\xi_2, f_3(x) = 2\xi_1 - \xi_2\}$.

Один из двух образов, на которые комитет делит пространство R^2 , показан на рис.11 штриховкой.

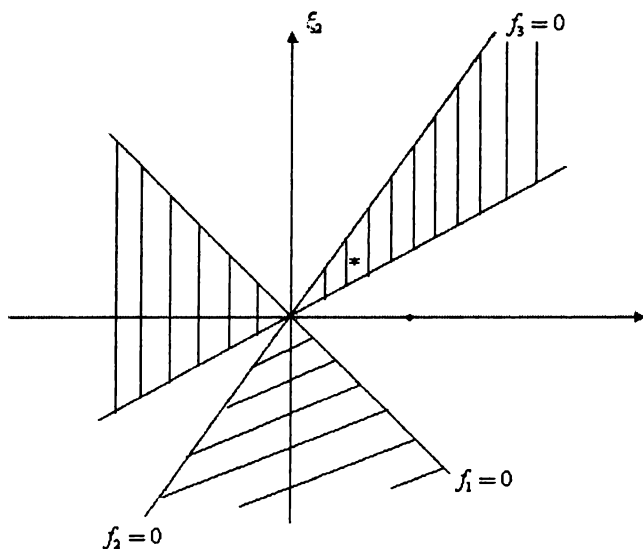


Рис.11.

1.4. Комитеты в распознавании образов

Рассмотрим задачу дискриминантного анализа (1). Пусть система (1) несовместна, но существует её комитет $K \subset F$. Тогда можно разбить R^n на два образа согласно следующему решающему правилу:

$$y - \text{элемент} \begin{cases} \text{первого образа, если } f(y) > 0, \text{ более,} \\ \text{чем для половины элементов } f \in K; \\ \text{иначе второго образа.} \end{cases}$$

§ 2. Теоремы существования

2.1. Теорема существования p -комитета системы множеств

Если всякие k множеств конечного семейства \mathcal{D} множеств: $\mathcal{D} = \{D_1, \dots, D_m\}$ имеют непустое пересечение, то при $k/m > p$ система \mathcal{D} обладает p -комитетом.

Доказательство. Пусть $A = \{A_1, \dots, A_t\}$ - семейство всех пересечений вида $D_{i_1} \cap \dots \cap D_{i_k}, i_t \neq i_u (t \neq u)$. Пусть K - система различных представителей для семейства A . Тогда K есть p -комитет семейства \mathcal{D} , так как $\forall j: |K \cap D_j|/|K| \geq C_m^{t-1}/C_m^t = k/m > p$, что и требовалось.

2.2. Условия существования комитета линейных неравенств

Рассмотрим конечную систему линейных неравенств относительно $x \in R^n$:

$$\langle c_j, x \rangle > 0 \quad (\forall j \in \overline{1, m}). \quad (5)$$

Теорема. Для существования комитета системы (5) необходимо и достаточно выполнения следующих двух условий:

а) Нет нулевых векторов: $c_j \neq 0 \quad (\forall j \in \overline{1, m})$, где 0 - нулевой вектор;

б) Никакие два вектора c_i и c_j при $i \neq j$ не являются противоположно направленными друг к другу.

Замечание. Условия а) и б) вместе равносильны следующему: если $\lambda \geq 0$ и $i \neq j$, то

$$c_i = -\lambda c_j. \quad (6)$$

Систему линейных однородных неравенств, удовлетворяющую условию (6), назовем системой без противоречий.

Доказательство теоремы.

1. *Необходимость.* Пусть система (5) обладает комитетом $k \in R^n$. Докажем, что в этом случае выполняются условия а) и б). Доказательство проведем методом от противного: пусть хотя бы одно условие а) или б) не выполняется. Пусть, например, не выполняется а). Тогда $\exists j: c_j = 0$ и потому $\langle c_j, x \rangle = 0 \ (\forall x \in K)$, что противоречит определению комитета. Пусть не выполняется б), т.е. для некоторых номеров i и j и для некоторого $\lambda > 0: c_i = -\lambda c_j$.

По определению комитета,

$$|\{x \in K: \langle c_i, x \rangle > 0\}| > |\{x \in K: \langle c_j, x \rangle \leq 0\}|;$$

$$|\{x \in K: \langle c_j, x \rangle > 0\}| > |\{x \in K: \langle c_j, x \rangle \leq 0\}|.$$

$$\text{Тогда } \{x \in K: \langle c_i, x \rangle > 0\} \cap \{x \in K: \langle c_j, x \rangle > 0\} = \emptyset.$$

Действительно, в противоположном случае было бы:

$$K / \{x \in K: \langle c_i, x \rangle > 0\} \supset \{x \in K: \langle c_j, x \rangle > 0\},$$

$$K / \{x \in K: \langle c_j, x \rangle > 0\} \supset \{x \in K: \langle c_i, x \rangle > 0\},$$

т.е.

$$\{x \in K: \langle c_i, x \rangle \leq 0\} \supset \{x \in K: \langle c_j, x \rangle > 0\},$$

$$\{x \in K: \langle c_j, x \rangle \leq 0\} \supset \{x \in K: \langle c_i, x \rangle > 0\},$$

откуда

$$|\{x \in K: \langle c_i, x \rangle \leq 0\}| \geq |\{x \in K: \langle c_j, x \rangle > 0\}| > |\{x \in K: \langle c_j, x \rangle \leq 0\}| \geq |\{x \in K: \langle c_i, x \rangle > 0\}| >$$

$$> |\{x \in K: \langle c_i, x \rangle \leq 0\}| - \text{противоречие.}$$

Итак, $\exists x \in K: \langle c_i, x \rangle > 0, \langle c_j, x \rangle > 0$. Но ввиду $c_i = -\lambda c_j, \lambda > 0$, это невозможно. Полученное противоречие окончательно доказывает необходимость.

2. *Достаточность.* Пусть $c_i \neq -\lambda c_j$ при $\lambda \geq 0$ и $i \neq j$. Не ограничивая общности, мы можем предполагать, что при $i \neq j$ и $\lambda > 0: c_i \neq -\lambda c_j$, так как в противном случае из двух неравенств $\langle c_i, x \rangle > 0$ и $\langle c_j, x \rangle > 0$ одно можно было бы удалить из системы (5), не нарушив хода доказательства. Итак, мы предполагаем, что $c_i \neq -\lambda c_j$ при $i \neq j$ и любом числе λ .

Докажем, что в этом случае

$$\forall i \in \overline{1, m}, \exists x' \in R^n : \langle c_i, x' \rangle = 0, \langle c_j, x' \rangle \neq 0, (i \neq j) \quad (7)$$

Действительно, $\exists i \in \overline{1, m}$, если $\langle c_i, x \rangle = 0$, то $j \neq i: \langle c_j, x \rangle = 0$. Пусть, например, если $\langle c_1, x \rangle = 0$, то $\exists i \in \overline{2, m}: \langle c_i, x \rangle = 0$. Положим $H_j = \{x: \langle c_j, x \rangle = 0\}$. По нашему предположению, объединение конечного числа гиперплоскостей $\bigcup_{j=2}^m H_j$ полностью содержит гиперплоскость H_1 , не совпадающую ни с одной из них, что невозможно.

Итак, мы доказали утверждение (7). Построим множество

$$K = \{x' + \varepsilon c_1; -x' + \varepsilon c_1; x^2 + \varepsilon c_2; -x^2 + \varepsilon c_2; \dots; x^m + \varepsilon c_m; -x^m + \varepsilon c_m\},$$

где число ε удовлетворяет условию

$$0 < \varepsilon < \frac{\min_{i \neq j} |\langle x', c_i \rangle|}{\max_{i \neq j} |\langle c_i, c_j \rangle|}.$$

Легко видеть, что множество K – комитет системы (5). Действительно, рассмотрим i -е неравенство этой системы: $\langle c_i, x \rangle > 0$.

Ему удовлетворяют векторы $x' + \varepsilon c_i; -x' + \varepsilon c_i$, а из каждой пары $x' + \varepsilon c_j; -x' + \varepsilon c_j$ ($j \neq i$) ему удовлетворяет хотя бы один

вектор. Следовательно, i -му неравенству удовлетворяют $2 + (m - 1) = (m + 1)$ векторов множества K - больше половины всех $2m$ членов этого множества.

Итак, теорема полностью доказана.

Рассмотрим неоднородную систему линейных неравенств

$$\langle c_j, x \rangle > \alpha_j \quad (\forall j \in \overline{1, m}) \quad (8)$$

Теорема. Если $[c_j = 0] \Rightarrow [\alpha_j < 0]$ и $[c_i \neq 0, c_j \neq 0, i \neq j] \Rightarrow \frac{c_i}{\|c_i\|} \neq -\frac{c_j}{\|c_j\|}$, то комитет системы (8) существует.

Доказательство. Ввиду условия теоремы, можно предполагать, что $c_j = 0 \ (\forall j)$; тогда выполнено условие (7).

Следовательно, существует некоторый комитет $K = \{x^1, \dots, x^q\}$ системы (5). Так как при достаточно большом $\lambda > 0$: $[\langle c_j, x' \rangle > 0] \Rightarrow [\langle c_j, \lambda x' \rangle > \lambda_j]$, то множество $\lambda K = \{\lambda x^1, \dots, \lambda x^q\}$ является комитетом системы (8).

Теорема доказана.

2.3. Оценка числа членов минимального комитета

Определение 1. Минимальным называется комитет с наименьшим возможным числом членов.

Определение 2. Максимальной (по включению) совместной подсистемой для системы неравенств называется такая её совместная подсистема, которая становится несовместной при добавлении любого нового неравенства системы. Максимальные совместные подсистемы будем кратко называть μ -подсистемами.

Теорема. Если комитет системы (5) существует, то существует и минимальный комитет с числом членов $q_{\min} \leq m$. Эта оценка является точной.

Доказательство. Так как комитет существует, то выполняется условие (7). Не ограничивая общности, будем предполагать, что среди c_j нет и одинаково направленных векторов, т.е.

$$\forall \alpha \in R^1 : c_j + \alpha c_j \neq 0 \ (i \neq j). \quad (9)$$

Покажем, что существует $(n \times 2)$ - матрица A , для которой

$$\forall \alpha \in R^1 : c_j A + \alpha c_j A \neq 0 = (0, 0).$$

Обозначим столбцы матрицы A через $u \in R^n$ и $v \in R^2$:

$$A = (u; v); c_j A = (\langle c_j, u \rangle, \langle c_j, v \rangle) \in R^2$$

В качестве u выберем такой вектор, что $\langle c_j, u \rangle \neq 0 \ (\forall j)$.

Введем обозначение: $\alpha_j = \langle c_j, u \rangle / \langle c_j, u \rangle$. По условию (9), $c_j - \alpha_j c_j \neq 0$.

Тогда $\exists v \in R^2 : \langle c_j - \alpha_j c_j, v \rangle \neq 0 \ (i \neq j)$.

Найденные векторы u и v и образуют, очевидно, столбцы искомой матрицы A .

Системе (5) поставим в соответствие систему

$$\langle c_j A, y \rangle > 0 \ (\forall j \in \overline{1, m}), y \in R^2 \quad (10).$$

Мы показали, что если система (5) без противоречий, то существует линейное преобразование A такое, что система (10) также без противоречий.

Заметим, что если $K = \{y_1, \dots, y_q\}$ - комитет системы (10), то множество

$\{y_1 A^T, \dots, y_q A^T\}$, где A^T - матрица, транспонированная по отношению к A , является комитетом системы (5), так как $\langle c_j A, y' \rangle = \langle c_j, y_i A^T \rangle$.

Следовательно, нам достаточно доказать, что если (5) - система над R^2 без противоречий, то существует её комитет с числом членов $q \leq m$.

Итак, предположим, что (5) - система без противоречий над пространством R^2 . Покажем, что тогда, если для каждой максимальной совместной подсистемы (μ - подсистемы) системы (5) взять по одному решению, то полученное множество, будет комитетом системы (5), причём число всех

μ - подсистем (а следовательно, и число членов этого комитета) не больше числа всех неравенств системы.

Точнее, покажем, что при $n = 2$ число q всех μ -подсистем системы без противоречий (5) нечётно: $q = 2k + 1$, причём $q \leq m$, и если мы произвольно зафиксируем номер $j' \in \overline{1, m}$, то число μ -подсистем, множества решений которых лежат в полуплоскости решений j' -го неравенства, равно $k + 1$.

Действительно, легко проверить, что число тех μ -подсистем системы (5), множество решений которых лежат в полуплоскости $\{x: \langle c_{j'}, x \rangle > 0\}$, равно числу μ -подсистем системы

$$\left. \begin{aligned} \langle c_{j'}, x \rangle &= 1, \\ \langle c_j, x \rangle &> 0 \quad (\forall j \in \overline{1, m} / \{j'\}), \end{aligned} \right\} \quad (11)$$

включающих уравнение $\langle c_{j'}, x \rangle = 1$. В то же время соответствующее число для полуплоскости $\{x: \langle c_{j'}, x \rangle < 0\}$ равно числу тех μ -подсистем системы

$$\left. \begin{aligned} \langle c_{j'}, x \rangle &= -1, \\ \langle c_j, x \rangle &> 0 \quad (\forall j \in \overline{1, m} / \{j'\}), \end{aligned} \right\} \quad (12)$$

включающих уравнение $\langle c_{j'}, x \rangle = -1$, множества решений которых ограничены.

Так как система (11) в качестве множества решений своих μ -подсистем имеет некоторое множество интервалов на прямой, а системе (12) соответствует система внутренних дополнений этих интервалов, то утверждение, что $q = 2k + 1$ число μ -подсистем множества решений которых лежат в j' -ой полуплоскости, равно $k + 1$, доказано. Тем самым доказана и теорема.

Замечание. Оценка этой теоремы является точной в следующем смысле: для всякого нечётного m существует

система m линейных неравенств вида (5), минимальный комитет которой состоит из m членов. Соответствующие примеры легко строятся для случая R^2 . В то же время для систем большой размерности (т.е. для больших m и n) обычно число членов минимального комитета много меньше, чем m . В частности, легко видеть, что минимальный комитет всякой неприводимо несовместной системы линейных неравенств (т.е. такой несовместной системы, всякая собственная подсистема которой совместна) состоит из трёх членов.

2.4. Существование разделяющих комитетов

Теорема. Пусть $A, B \in R^n$. Для того чтобы существовал разделяющий A и B линейный комитет, необходимо и достаточно, чтобы

- а) $0 \notin A \cup B$;
- б) каждое из множеств A и B было множеством без противоположно направленных векторов;
- в) если $a \in A$ и $b \in B$, то a и b не являлись бы одинаково направленными векторами.

Доказательство получается непосредственным применением теоремы о существовании комитета системы (5).

Теорема. Пусть $A, B \in R^n$. Для того, чтобы существовал разделяющий множества A и B комитет аффинных функций (т.е. аффинный комитет) необходимо и достаточно выполнения условия $A \cap B = \emptyset$.

Доказательство. Легко убедиться, что при $A \cap B = \emptyset$ система

$$\left. \begin{aligned} \langle a, x \rangle + y &> 0 \quad (\forall a \in A), \\ -\langle b, x \rangle + (-y) &> 0 \quad (\forall b \in B) \end{aligned} \right\}$$

рассматривается относительно $x \in R^n, y \in R^l$, является системой без противоречий. Отсюда и следует утверждение теоремы.

§ 3. Алгоритмы построения комитетов

3.1. Алгоритм с дообучением

Опишем алгоритм, который соответствует доказательству теоремы об оценке числа членов минимального комитета. Это один из простых алгоритмов нахождения комитета и его преобразования при добавлении новых неравенств, что соответствует предъявлению новых объектов материала обучения.

Пусть задана система неравенств (она может быть потенциально бесконечной, т.е. иметь вид последовательности неравенств, но для удобства изложения мы примем, что она конечна с достаточно большим числом m неравенств):

$$\langle c_j, x \rangle > 0 \quad (\forall j \in \overline{1, s}) \quad (1)$$

Будем предполагать, что (1) – система без противоречий.

Введем обозначения:

$$c_j = (\alpha_{j_1}, \dots, \alpha_{j_n}),$$

$$u = (\eta_1, \dots, \eta_n),$$

$$v = (\omega_1, \dots, \omega_n).$$

Пусть вначале задано небольшое число s элементов обучения, т.е. c_1, \dots, c_s . При выбранных u и v предлагаемый алгоритм построения и преобразования комитета однородной системы

$$\langle c_j, x \rangle > 0 \quad (\forall j \in \overline{1, s}) \quad (2)$$

состоит в следующем.

1. Вычисляются двумерные векторы

$$d_j = \left(\alpha_{j,n-1} + \sum_{k=1}^{n-2} \alpha_{jk} \eta_k, \alpha_{j,n} + \sum_{k=1}^{n-2} \alpha_{jk} \omega_k \right) \quad (j = 1, \dots, s).$$

Предполагается: u и v таковы, что векторы d_j образуют систему без противоречий.

2. Находится комитет системы линейных неравенств над пространством R^2

$$\langle d_j, y \rangle > 0 \quad (\forall j \in \overline{1, s}), \quad (3)$$

Состоящей из решений её μ -подсистем.

Пусть $\{y_1, \dots, y_q\}$ - комитет системы (3), $y_i = (\xi'_i, \xi''_i)$. Тогда легко видеть, что множество $\{x_1, \dots, x_q\}$, где $x_i = (\xi'_i, \xi''_i)$,

$$\xi'_k = \eta_k \xi'_{n-1} + \omega_k \xi'_n \quad (k = 1, \dots, n-1, i = 1, \dots, q) \quad (4)$$

является комитетом системы (2).

3. Если в системе (2) добавляется новое неравенство $\langle c_{s+1}, x \rangle > 0$, то образуем вектор

$d_{s+1} = \left(\alpha_{s+1,n-1} + \sum_{k=1}^{n-2} \alpha_{s+1,k} \eta_k, \alpha_{s+1,n} + \sum_{k=1}^{n-2} \alpha_{s+1,k} \omega_k \right) = (\delta_{s+1}, \delta_{s+1,2})$ и переходим к следующему шагу.

4. При добавлении неравенства $\langle d_{s+1}, y \rangle > 0$ к системе (3) возможны только следующие три случая:

а) множество $\{y_1, \dots, y_q\}$, являясь комитетом системы (3), является комитетом и системы

б)

$$\langle d_j, y \rangle > 0 \quad (\forall j \in \overline{1, s+1}). \quad (5)$$

В этом случае преобразовывать его нет необходимости.

в) множество $\{y_1, \dots, y_q\}$ не является комитетом системы (5), и хотя бы для одного элемента $y_{i_0}, i_0 \in \overline{1, q}$, некоторая точка z полупространства

$$\{y: \langle d_{s+1}, y \rangle > 0\}$$

принадлежат той же μ -подсистеме системы (2), что и y_{i_0} .

Тогда при условии, что вектор d_{s+1} не является противоположно направленным по отношению ни к одному вектору d_1, \dots, d_s , множество

$$\{y: i \neq i_0\} \cup \{z\}$$

является комитетом системы (5);

г) в последнем из возможных случаев при условии, что система (5) – без противоречий, множество

$\{y_1, \dots, y_q; (-\delta_{s+1,2}; \delta_{s+1,1}) + \varepsilon d_{s+1}; (\delta_{s+1,2}; -\delta_{s+1,1}) + \varepsilon d_{s+1}\}$ есть комитет системы (5) при достаточно малом $\varepsilon > 0$.

5. По формулам (4) вычисляются новые члены комитета системы $\langle c_j, x \rangle > 0 \ (\forall j \in \overline{1, s+1})$.

3.2. Алгоритм линейной коррекции

Рассмотрим систему (1). Из множества $\{c_j\}_{j=1}^m$ составляется бесконечная последовательность $\{g_j\}_{j=1}^{+\infty}$, в которой множество $\{c_j\}_{j=1}^m$ циклически повторяется: $g_1 = c_1, \dots, g_m = c_m, g_{m+1} = c_1, \dots, g_{2m} = c_m, g_{2m+1} = c_1, \dots$. Комитет строится в результате последовательности приближений. Вначале выбирается произвольное множество $K_0 = \{x_1(0), \dots, x_q(0)\}$, число q фиксировано заранее из тех или иных эвристических

соображений. Переход от i -го приближения к $(i+1)$ -му ($i = 0, 1, 2, \dots$) производится согласно следующему правилу.

Пусть $K_0 = \{x_1(i), \dots, x_q(i)\}$.

Вычисляем $N_i = \sum_{j=1}^q \text{sign} \langle x_j(i), g_{i+1} \rangle$.

Затем выбираем $\frac{1}{2}(|N_i|+1)$ векторов $x_j(i)$, имеющих наименьшие по абсолютной величине отрицательные скалярные произведения с g_{i+1} . Каждый из этих векторов изменяется по правилу $x_j(i+1) = x_j(i) + g_{i+1}$.

Этот метод не всегда приводит к нахождению комитета.

§ 4. Построение минимального комитета

4.1. Метод свёртывания

Метод свёртывания (исключения неизвестных) для систем линейных неравенств, в том числе для несовместных систем, разработан и исследован С. Н. Черниковым. Приводимые в настоящем пункте 4.1. результаты принадлежат С. Н. Черникову.

Рассмотрим систему линейных неравенств

$$\left. \begin{aligned} \langle c_j, x \rangle - \beta_j &\equiv \alpha_{j1}\xi_1 + \dots + \alpha_{jm}\xi_m - \beta_j < 0 \quad (j=1, \dots, m') \\ \langle c_j, x \rangle - \beta_j &\equiv \alpha_{j1}\xi_1 + \dots + \alpha_{jm}\xi_m - \beta_j \leq 0 \quad (j=m'+1, \dots, m) \end{aligned} \right\} \quad (1)$$

Определение 1. Пусть U - подпространство из R^n ;
 $C(U) = \left\{ u = (\eta_1, \dots, \eta_m) \geq 0_m : \sum_{j=1}^m \eta_j \langle c_j, x \rangle = 0 (\forall x \in U) \right\}; \{ y_i = (\eta_{i1}, \dots, \eta_{im}) : i = 1, \dots, k \}$
- некоторая конечная система порождающих элементов конуса $C(U)$.

Элементу y_i ставим в соответствие неравенство

$$\sum_{j=1}^m \eta_{ij} \langle c_j, x \rangle - \sum_{j=1}^m \eta_{ij} \beta_j \leq 0,$$

если $m' = 0$ или $\eta_{i1} = \dots = \eta_{im'} = 0$ при $m' > 0$ и неравенство

$$\sum_{j=1}^m \eta_{ij} \langle c_j, x \rangle - \sum_{j=1}^m \eta_{ij} \beta_j < 0,$$

если $m' > 0$ и хотя бы один из коэффициентов $\eta_{i1}, \dots, \eta_{im}$ отличен от нуля.

Система всех полученных таким образом неравенств называется U -свёрткой системы (1). Если $C(U) = \{ o = (0, \dots, 0) \}$, то U -свёртка называется пустой.

Если $\{y_i, i=1, \dots, k\}$ - базис конуса $C(u)$, то u -свёртка называется фундаментальной. При $U = R^n : U$ -свёртка называется полной. Индексом ненулевого элемента конуса $C(u)$ называется совокупность номеров ненулевых координат этого элемента. Индексами неравенств U -свёртки системы (1) называются индексы соответствующих им элементов конуса $C(u)$.

Определение 2. Ненулевой элемент конуса $C(u)$, который имеет положительные координаты с номерами j_1, \dots, j_{i+1} , и нулевые остальные координаты, называются фундаментальным элементом этого конуса, если s функций из $\langle c_{j_1}, x \rangle, \dots, \langle c_{j_{i+1}}, x \rangle$ линейно независимы на U .

Теорема 1. Конус $C(u)$ имеет конечное множество существенно различных фундаментальных элементов. Максимальные системы существенно различных фундаментальных элементов конуса $C(u)$, и только они суть его базисы. (Два элемента конуса $C(u)$ не являются существенно различными, если, и только если, они отличаются друг от друга только положительным множителем).

Следствие. Не пустая U -свёртка системы (1) является фундаментальной тогда и только тогда, когда она не содержит ни одного неравенства с индексом, включающим индекс хотя бы одного неравенства с индексом, включающим индекс хотя бы одного из остальных её неравенств.

Определение 3. Пусть U и U' - два подпространства.

Если U -свёртка s системы (1) не пуста, то U' -свёртка системы s называется повторной $(U; U')$ -свёрткой системы (1). Если система s пуста или её U' -свёртка пуста, то повторная $(U; U')$ -свёртка системы (1) пуста. Аналогично определяется повторная $(U; U'; U'')$ -свёртка системы (1) и т.д.

Теорема 2. Любая $(U; U')$ -свёртка системы (1) совпадает с некоторой $(U + U')$ -свёрткой этой системы.

Теорема 3. Если система (1) совместна, то для всякого подпространства $U \subset R^n : U$ - свёртка системы (1) совместна или пуста. Если существует подпространство $U \subset R^n$ и U -свёртка системы (1), совместная или пустая, то система (1) совместна.

Определение 4. Пусть $A = \{\alpha_1, \dots, \alpha_m\}$. Каждой паре $\alpha_p > 0, \alpha_q < 0$ соотнесём неравенство $\alpha_p(\langle c_q, x \rangle - \beta_q) - \alpha_q(\langle c_p, x \rangle - \beta_p) < 0$, если $m' > 0$ и хотя бы один из номеров p, q не больше m' , или неравенство $\alpha_p(\langle c_q, x \rangle - \beta_q) - \alpha_q(\langle c_p, x \rangle - \beta_p) \leq 0$ в противном случае. Каждому элементу $\alpha_s \in A$ отнесем неравенство $\langle c_s, x \rangle - \beta_s < 0$, если $m' > 0$ и $s \leq m'$, или неравенство $\langle c_s, x \rangle - \beta_s \leq 0$ в противном случае. Совокупность S всех таких неравенств называется полученной в результате A -деформации системы (1).

Теорема 4. Если $x_0 \in R^n, U(x_0)$ - порожаемое элементом x_0 подпространство, $A(x_0) = \{\alpha_j = \langle c_j, x_0 \rangle : j = 1, \dots, m\}$, то $A(x_0)$ - деформация системы (1) есть $U(x_0)$ - свертка системы (1).

4.2. Алгоритм фундаментального свертывания

Пусть $U_i = U(e_i), e_i$ - i -й координатный орт. Алгоритм состоит из следующих шагов.

а) Пусть A_i - столбец коэффициентов при ξ_i . Произведя A_i -деформацию системы (1), получим систему S_i , являющуюся фундаментальной U_i - сверткой системы (2). Каждому неравенству системы S_i отнесем его индекс.

б) Пусть уже получена система S_k - фундаментальная $(U_1 + \dots + U_k)$ - свертка системы (1). Пусть A_{k+1} - столбец коэффициентов при ξ_{k+1} . Произведем A_{k+1} -деформацию системы S_k , не комбинируя тех пар неравенств, объединение индексов которых содержит индекс какого-либо третьего неравенства системы S_k . Получим систему S_{k+1} -

фундаментальную $(U_1 + \dots + U_{k+1})$ -свертку системы (1). Каждому её неравенству отнесем его индекс.

в) Если $S_i = \emptyset$, то полагаем $S_{i+1} = \emptyset$. Через n шагов получим полную фундаментальную свертку системы (1).

Теорема 5. Номера неравенств несовместной системы (1), которые входят в её произвольную минимальную несовместную подсистему, образует индекс одного из несовместных неравенств полной фундаментальной свертки системы (1). Индекс каждого из несовместных неравенств упомянутой свертки состоит из номеров неравенств, образующих некоторую минимальную несовместную подсистему системы (1).

4.3. Нахождение минимального комитета

Рассмотрим систему над R^n :

$$\langle c_j, x \rangle > 0 \quad (\forall j \in \overline{1, m}) \quad (2)$$

Пусть имеется некоторая подсистема системы (2):

$$\langle c_j, x \rangle > 0 \quad (\forall j \in J), \quad J \subset \overline{1, m}. \quad (3)$$

Множество J будем называть индексом системы (3). Максимальная (по включению) совместная подсистема (т.е. μ -подсистема) системы (2) есть такая совместная подсистема (3), что если к ней добавить любое новое неравенство из системы (2), то она станет несовместной. Минимальная (по включению) несовместная подсистема (т.е. ν -подсистема) системы (2) есть такая несовместная подсистема, что если из неё удалить любое её неравенство, то получим совместную подсистему.

Теорема. Среди минимальных комитетов системы (2), т.е. комитетов с наименьшим возможным числом членов, существует такой, все члены которого суть решения некоторых μ -подсистем системы (2).

Доказательство. Пусть множество $K = \{x_1, \dots, x_q\}$ есть минимальный комитет системы (2). Определим для всякого $i \in \overline{1, q}$ множество $J_i = \{j \in \overline{1, m}, \langle c_j, x_i \rangle > 0\}$.

Рассмотрим систему $\langle c_j, x \rangle > 0 (\forall j \in J_i)$. Так как эта система совместна, то существует μ -подсистема, индекс которой включает J_i ; пусть y_i - некоторое решение этой μ -подсистемы. Легко видеть, что множество $\{y_1, \dots, y_q\}$ является минимальным комитетом системы (2), все члены которого суть решения некоторых μ -подсистем.

Теорема доказана.

Эта теорема позволяет свести задачу нахождения минимального комитета системы (2) к двум следующим.

1. Найти все μ -подсистемы системы (2) и для каждой μ -подсистемы найти её решение; пусть эти решения составляют множество $\{x_1, \dots, x_r\}$.

2. Определить, сколько экземпляров каждого элемента x_i должно войти в минимальный комитет.

Задачу 2 можно решить с помощью методов целочисленного программирования. Предположим, что множество $\{x_1, \dots, x_r\}$ уже найдено и определим $S_i = \{j : \langle c_j, x_i \rangle > 0\} (i = 1, \dots, r)$.

Каждому S_i поставим в соответствие вектор $\omega_i = (\omega_{i1}, \dots, \omega_{im})$, где $\omega_{ij} = \begin{cases} +I, & \text{если } j \in S_i; \\ -I, & \text{если } j \notin S_i. \end{cases}$

Пусть λ_i - искомое число экземпляров элемента x_i , которые должны войти в минимальный комитет. Тогда задача нахождения минимального комитета сводится к следующей задаче целочисленного программирования: минимизировать $(\lambda_1 + \dots + \lambda_r)$ при условиях

$$\begin{cases} \lambda_1 \omega_1 + \dots + \lambda_r \omega_r \geq (1, \dots, 1), \\ \lambda_i \in \{0, 1, 2, \dots\} \quad (\forall i \in \overline{1, r}). \end{cases}$$

Перейдем к задаче 1. Покажем, что все μ -подсистемы системы (2) можно найти зная все ν -подсистемы (т.е. минимальные несовместные). Действительно, пусть t_1, \dots, t_r - индексы всех ν -подсистем системы (2). Тогда ясно, что множество s является индексом совместной подсистемы в том и только том случае, если S не содержит T_i .

Поэтому индексы μ -подсистем можно определить как максимальные по включению s , удовлетворяющие этому условию.

Итак, задача свелась к нахождению всех ν -подсистем. Эти подсистемы можно найти методом свертывания как следует из теоремы 5 (С. Н. Черникова), приведенной в 4.1.

Рассмотрим пример. Пусть требуется найти минимальный комитет следующей системы линейных неравенств:

$$\left. \begin{array}{l} (1) \quad \xi_1 > 0, \\ (2) \quad \xi_2 > 0, \\ (3) - \xi_1 - \xi_2 > 0, \\ (4) - \xi_1 + 2\xi_2 > 0, \\ (5) 2\xi_1 - \xi_2 > 0. \end{array} \right\}$$

Слева от каждого неравенства записан его индекс. Найдем фундаментальную ξ_1 -свертку:

$$\left. \begin{array}{l} (1,3) \quad -\xi_2 > 0, \\ (1,4) \quad 2\xi_2 > 0, \\ (2) \quad \xi_2 > 0, \\ (3,5) \quad -3\xi_1 > 0, \\ (4,5) \quad 3\xi_1 > 0. \end{array} \right\}$$

Фундаментальная ξ_2 -свертка последней системы есть полная фундаментальная свертка исходной системы:

$$(1, 3, 4) \quad 0 > 0,$$

$$(1, 2, 3) \quad 0 > 0,$$

$$(2, 3, 5) \quad 0 > 0,$$

$$(3, 4, 5) \quad 0 > 0.$$

Индексы несовместных неравенств (т.е. неравенств вида $0 > 0$) полной фундаментальной свертки суть индексы ν -подсистем исходной системы:

$$T_1 = \{1, 3, 4\}, T_2 = \{1, 2, 3\}, T_3 = \{2, 3, 5\}, T_4 = \{3, 4, 5\}.$$

Индексы S_i μ -подсистем – максимальные $S_i \subset \overline{1, 5}$, $i = 1, 2, 3$, удовлетворяющие условию: $S \not\supset T_i (\forall i \in \overline{1, 4})$:

$$S_1 = \{1, 2, 4, 5\}, S_2 = \{2, 3, 4\}, S_3 = \{1, 3, 5\}.$$

Для каждого $i \in \overline{1, 3}$ найдем x_i -решение i -й μ -подсистемы:

$$x_1 = (1, 1); x_2 = (-2, 1); x_3 = (1, -2).$$

Определим векторы $\omega_1, \omega_2, \omega_3$:

$$\omega_1 = (1, 1, -1, 1, 1), \omega_2 = (-1, 1, 1, 1, -1), \omega_3 = (1, -1, 1, -1, 1).$$

Составляем задачу целочисленного программирования:

$$\min \{ \lambda_1 + \lambda_2 + \lambda_3 : \lambda_1 \omega_1 + \lambda_2 \omega_2 + \lambda_3 \omega_3 \geq (1, 1, 1, 1, 1), \lambda_i \in \{0, 1, 2, \dots\} (\forall i \in \overline{1, 3}) \};$$

Она расписывается в виде

$$\min(\lambda_1 + \lambda_2 + \lambda_3),$$

$$\lambda_1 - \lambda_2 + \lambda_3 \geq 1;$$

$$\lambda_1 + \lambda_2 - \lambda_3 \geq 1;$$

$$-\lambda_1 + \lambda_2 + \lambda_3 \geq 1;$$

$$\lambda_1 + \lambda_2 - \lambda_3 \geq 1;$$

$$\lambda_1 - \lambda_2 + \lambda_3 \geq 1;$$

$$\lambda_i \in \{0, 1, 2, \dots\} \quad (i = 1, 2, 3)$$

Решение последней задачи: $\lambda_1 = \lambda_2 = \lambda_3 = 1$.

Итак, один из минимальных комитетов имеет вид:

$$K = \{x_1, x_2, x_3\}.$$

§ 5. Вопросы программного обеспечения

5.1. Необходимость создания пакетов программ распознавания образов

Последнее время широкое распространение получил новый подход к созданию высокоэффективных программных систем обработки данных; он состоит в разработке пакетов прикладных программ. Сущность этого подхода в том, что большие программы и комплексы программ рассматриваются как структуризованные системы, строящиеся из элементарных программных единиц, называемых модулями.

Необходимость такого подхода диктуется современным этапом развития математической технологии решения больших прикладных задач, требованиями к качеству, времени и надёжности решения этих задач. Н. Н. Яненко отмечает внутренние и внешние факторы, требующие структуризации больших программ:

Внутренние факторы программирования: требование иерархического разбиения всего объёма работ по программированию, их контролируемости, требование надёжности.

Среди внешних факторов отмечаются: наличие структуры в самих вычислительных алгоритмах, необходимость обмена программами, необходимость эволюции программ.

Так как методы распознавания образов нацелены на решение прикладных задач классификации, диагностики и управления, проблема создания пакетов прикладных программ является актуальной и для предметной области. Теоретические основы структуризации классов алгоритмов, распознавания и классификации, необходимой при создании структуризованных больших программ-пакетов, разработаны Ю. И. Журавлёвым. В настоящее время создан ряд пакетов

программ распознавания и классификации. Коротко опишем один из них.

5.2. Пакет «КВАЗАР»

Пакет прикладных программ распознавания образов «КВАЗАР» разработан в Институте математики и механики УНЦ АН СССР В.С. Казанцевым при участии других сотрудников института. В основе алгоритмического наполнения пакета лежат комитетные конструкции для систем линейных неравенств.

Пакет представляет собой систему программ на алгоритмическом языке ФОРТРАН - ГДР, реализующих алгоритмы решения задач таксономии (обучение ЭВМ без учителя) дискриминантного анализа (обучение с учителем), отображения многомерных данных в пространство меньшей размерности (в том числе в двумерное), выбора информативного признакового пространства и некоторых других задач.

Пакет написан для ЭВМ БЭСМ-6. В нём содержится библиотека программ-модулей, в которых реализованы некоторые алгоритмы и процедуры, а также головная программа, имеющая блок трансляции заданий и блок управления решением задачи. Пакет «КВАЗАР» – открытая система: он с течением времени пополняется новыми модулями.

С помощью пакета «КВАЗАР» решены многие практические задачи экономики, техники, социологии, биологии и медицины.

Глава 4.

ОБЗОР МЕТОДОВ ДИСКРИМИНАНТНОГО АНАЛИЗА

К настоящему времени разработано множество методов дискриминантного анализа, их формализация и обоснование находятся на самых различных уровнях. Многие методы имеют лишь эвристическое основание, но оказались полезными при решении некоторых классов практических задач. Некоторые методы, ставшие стандартными, изучены достаточно глубоко; к ним относятся метод оценок Ю. И. Журавлёва, метод комитетов, метод потенциальных функций, методы стохастической аппроксимации и некоторые другие. Следует отметить работы Ю. И. Журавлёва, в которых изучается алгебраическая структура целых классов алгоритмов. В настоящей главе кратко рассматриваются некоторые из методов, ставших классическими и достаточно изученных с математической точки зрения.

§ 1. Метод потенциальных функций

1.1. Геометрический смысл метода

Метод изучается в книге [8].

Будем рассматривать детерминистскую постановку. Пусть $x, y \in R^n$. Через $K(x, y)$ будем обозначать потенциал, создаваемый точкой y в точках x . Функция $K(x, y)$ должна удовлетворять следующим условиям:

- а) $K(x, y) > 0 (\forall x, y \in R^n)$;
- б) $\max_{x \in R^n} K(x, y) = K(y, y)$;

в) если точка x удаляется от y , то значение $K(x, y)$ убывает.

Приведем примеры функции $K(x, y)$. Пусть $p(x, y) = \|x - y\|$. Тогда можно, например, положить $K = \exp(-\alpha p^2)$ или $K = (1 + \alpha p^2)^{-1}$, где $\alpha > 0$. График функции $K(x, y)$ при фиксированном y представляет собой «холм» с вершиной над точкой y .

Пусть $A, B \subset R^n$. Определим $K_A(x) = \sum_{y \in A} K(x, y)$ - сумму «холмов» над точками из A ; $K_B(x) = \sum_{y \in B} K(x, y)$. Тогда функцию $f(x) = K_A(x) - K_B(x)$ можно интерпретировать как разделяющую для множества A и B .

1.2. Итерационный процесс

Предположим, что искомая разделяющая функция $f(x)$ раскладывается в ряд по некоторой системе функций:

$$f(x) = \sum_i \alpha_i \varphi_i(x).$$

Строим приближения $f_k(x)$ к функции $f(x)$ согласно следующему правилу: $f_0(x) \equiv 0$; $f_{k+1}(x) = f_k(x) + r_k K(x, x_{k+1})$, где

$$r_k = \begin{cases} 0, & \text{если } f_k(x_{k+1}) \geq 0 \text{ и } x_{k+1} \in A; \\ 0, & \text{если } f_k(x_{k+1}) < 0 \text{ и } x_{k+1} \in B; \\ 1, & \text{если } f_k(x_{k+1}) < 0 \text{ и } x_{k+1} \in A; \\ -1, & \text{если } f_k(x_{k+1}) \geq 0 \text{ и } x_{k+1} \in B. \end{cases}$$

Теорема. (см. 8). Пусть $f(x) = \sum_{i=1}^{\infty} \alpha_i \varphi_i(x)$, $K(x, y) = \sum_{i=1}^{\infty} \alpha_i^2 \varphi_i(x) \varphi_i(y)$,

$\sum_{i=1}^{\infty} \left(\frac{\alpha_i}{\lambda} \right)^2 < +\infty$. Пусть далее существует $\varepsilon > 0$, для которого

$$f(x) \begin{cases} \geq \varepsilon, & x \in A; \\ \leq -\varepsilon, & x \in B. \end{cases}$$

Если точки x_i появляются независимо с некоторой плотностью $p(x)$, то математическое ожидание по x величины $|sign f(x) - sign f_k(x)|$ сходится к нулю почти наверное, при $k \rightarrow +\infty$.

§ 2. Методы, основанные на теории статистических решений

Образы – некоторые бесконечные множества в пространстве R^n . Относительно этих множеств могут быть известны те или иные вероятностные параметры или же эти параметры могут быть оценены по конечным выборкам из образов. Параметрические методы дискриминантного анализа:

- а) устанавливают зависимость дискриминантных функций от параметров;
- б) По материалу обучения оценивают параметры;
- в) Оценки подставляют вместо параметров в дискриминантные функции.

2.1. Задача разделения на k классов

Объекты классов появляются случайно, соответственно k функциям плотности распределения. В качестве параметров могут выступать:

$p(x/i)$ - вероятность появления x в предположении, что x принадлежит i -му образу;

$p(j)$ - вероятность появления элемента j -го образа;

$p(j/x)$ - вероятность, что x принадлежит j -му образу.

(Вместо «плотности распределения» будем для краткости писать «вероятность».)

Для разбиения R^n на k образов надо построить k функций $g_i(x)$ ($i=1, \dots, k$) таких, что если x принадлежит i -му образу, то $g_i(x) > g_j(x)$ ($i \neq j$).

2.2. Дискриминантные функции на основе теории решений

Пусть заданы потери $\lambda(i/j)$, когда объект j -го образа относят к i -му образу, и кроме того, параметры $p(j/x)$. Тогда условные средние потери от отнесения к i -му образу равны:

$$L_x(i) = \sum_{j=i, j \in 1, k} \lambda(i/j) p(j/x). \quad (1)$$

Поэтому правило отнесения x к одному из k образов таково: x относится к i_0 -му образу, где $L_x(i_0) = \min_i L_x(i)$.

Согласно формуле Байеса,

$$p(j/x) = p(x/j) p(j) / p(x). \quad (2)$$

Подставим выражение для $p(j/x)$ из (2) в (1), получим:

$$L_x(i) = \frac{1}{p(x)} \sum_{j=i, j \in 1, k} \lambda(i/j) p(x/j) p(j).$$

Для минимизации по i множитель $\frac{1}{p(x)}$ не имеет значения, поэтому вводим функции

$$\ell_x(i) = \sum_{j=i, j \in 1, k} \lambda(i/j) p(x/j) p(j). \quad (3)$$

Предположим далее, что функция является симметрической:

$$\lambda(i/j) = 1 - \delta_{ij}, \quad \delta_{ij} = \begin{cases} 1, & i = j, \\ 0, & i \neq j. \end{cases} \quad (4)$$

Подставим (4) в (3): $\ell_x(i) = p(x) - p(x/i) p(i)$.

Так как при фиксированном x : $p(x) = \text{const}$, то вместо минимизации $\ell_x(i)$ можно минимизировать по i функцию

$p(x/i)p(i)$. Прологарифмировав (что не влияет на максимизацию) получим i -ю дискриминантную функцию:

$$g_i(x) = \log p(x/i) + \log p(i) .$$

§ 3. Методы алгебры логики

3.1. Метод тупиковых тестов

Понятие теста было введено в 1956 г. С. В. Яблонским. На основе этого понятия Ю. И. Журавлёвым и другими разработаны тестовые алгоритмы распознавания образов, сыгравшие большую роль в разработке теории методов вычисления оценок.

Пусть материал обучения задан в виде таблицы T , строки которой – бинарные векторы состояния объектов, а столбцы – признаки:

Векторы	Признаки				Класс (образ)
	1	2	...	n	
x_1^1	ξ_{11}^1	ξ_{12}^1	...	ξ_{1n}^1	1
...
$x_{m_1}^1$	$\xi_{m_1 1}^1$	$\xi_{m_1 2}^1$...	$\xi_{m_1 n}^1$	1
...
x_1^i	ξ_{11}^i	ξ_{12}^i	...	ξ_{1n}^i	i
...
$x_{m_i}^i$	$\xi_{m_i 1}^i$	$\xi_{m_i 2}^i$...	$\xi_{m_i n}^i$	i
...
x_1^k	ξ_{11}^k	ξ_{12}^k	...	ξ_{1n}^k	k
...
$x_{m_k}^k$	$\xi_{m_k 1}^k$	$\xi_{m_k 2}^k$...	$\xi_{m_k n}^k$	k

Определение: Тестом таблицы T называется совокупность столбцов i_1, \dots, i_r таких, что после удаления из T всех столбцов, кроме столбцов с номерами i_1, \dots, i_r , получим

таблицу, в которой все пары строк, принадлежащих разным классам различны. Тест $\{i_1, \dots, i_r\}$ называется тупиковым, если никакая его истинная часть не является тестом.

Пусть T – множество всех тупиковых тестов таблицы T и $\Gamma \in T$, $\Gamma = \{i_1, \dots, i_r\}$. Классификация любого $y \in R^n$ проводится по следующему правилу.

Если $y = (\eta_1, \dots, \eta_n)$, то пусть $y(\Gamma) = (\eta_{i_1}, \dots, \eta_{i_r})$. Сравним $y(\Gamma)$ с векторами $x_1^1(\Gamma), \dots, x_m^1(\Gamma), \dots, x_1^t(\Gamma), \dots, x_m^t(\Gamma)$ таблицы T . Число совпадений $y(\Gamma)$ с векторами $x_i^1(\Gamma), \dots, x_m^1(\Gamma)$ i -го класса обозначим через $\gamma_i(\Gamma)$. Величина $\gamma'_i(\Gamma) = \frac{1}{m_i} \sum_{\Gamma \in T} \gamma_i(\Gamma)$ называется оценкой y по i -му классу. При этом y относится к классу i_0 , для которого $\gamma'_{i_0}(\Gamma) = \max_i \gamma'_i(\Gamma)$.

3.2. Метод анализа высказываний

Этот метод изучается в книге Горелика и Скрипкина [9]. Пусть имеются некоторые исходные высказывания A, B, C, \dots . Из них конструируются более сложные высказывания: $A + B$, $A \cdot B$, \bar{A} , $A \rightarrow B$, $A = B$ и т.д. истинным высказываниям присваивается значение 1, ложным – значение 0. Таблица, которая представляет все возможные комбинации значений истинности некоторого набора A, B, C, \dots , называется базисом.

Например, базис одного элемента:

N \ #	0	1
	0	1
# A	0	1

Базис двух элементов:

N \ #	0	1	2	3
# A	0	1	0	1
# B	0	0	1	1

Колонки базиса рассматриваются как двоичные числа (младший разряд – верхняя цифра), это числа от 0 до 2^{n-1} . Через #A обозначается изображающее число для A. Можно найти изображающее число любой булевой функции от A, B, ... по изображающим числам A, B, ...

Рассматриваются задачи распознавания образов, сводящиеся к задачам решения булевых уравнений такого, например, типа

$$X \cdot (A + B) = A \cdot B \cdot C \quad (*)$$

В качестве X отыскивается такая булевая функция от A, B, C, ..., чтобы подставив ей в (*), получить тавтологию. Решение получается на основе использования изображающих чисел.

3.3. Распознавание арифметических таблиц

В книге М. М. Бонгарда [10] рассматриваются, в частности, некоторые логические методы распознавания арифметических операций. Объект узнавания – таблица из нескольких строк. В каждой строке – три числа: $a; b; c = f(a, b)$, где f - некоторая арифметическая операция. Каждой операции f соответствует свой класс объектов (образ).

Следующим образом строятся признаки. Берутся несколько простых арифметических функций $f_i(a, b, c)$. Признаки имеют вид:

$$L_1 f_i(a, b, c) = \begin{cases} 1, & \text{если } f_i(a, b, c) - \text{целое число;} \\ 0, & \text{если } f_i(a, b, c) - \text{дробное число;} \end{cases}$$

$$L_2 f_i(a, b, c) = \begin{cases} 1, & \text{если } f_i(a, b, c) \geq 0; \\ 0, & \text{если } f_i(a, b, c) < 0; \end{cases}$$

$$L_3 f_i(a, b, c) = \begin{cases} 1, & \text{если } |f_i(a, b, c)| \geq 1; \\ 0, & \text{если } |f_i(a, b, c)| < 1. \end{cases}$$

Затем отбираются полезные признаки. Признак не должен присваивать всем таблицам одно число 0 или 1. таблицам одного класса признак должен приписывать одно и то же число.

Глава 5.

МЕТОДЫ ТАКСОНОМИИ

§ 1. Метод выделения максимальных совместных подсистем

Пусть $|X| \subset R^n$, $|X| < +\infty$. Пусть F - некоторое множество функций, $F \subset \{R^n \rightarrow R^1\}$. Будем считать допустимой формой таксона множество $\{x: f(x) \leq 0\}$ при некотором $f \in F$. Итак, таксоны множества X должны быть в виде: $X \cap \{x: f(x) \leq 0\}$.

Рассмотрим систему неравенств относительно неизвестной функции f :

$$f(x) \leq 0 \quad (\forall x \in X), f \in F. \quad (1)$$

Пусть все максимальные подсистемы (μ - подсистемы) системы (1) имеют вид: $f(x) \leq 0 \quad (\forall x \in X_j), f \in F, j = 1, \dots, k$. Иными словами, множества X_1, \dots, X_k суть все индексы μ - подсистем системы (1). Из совокупности $\{X_1, \dots, X_k\}$ надо выбрать разбиение $\{X_1, \dots, X_r\}$: $X = \bigcup_{j=1}^r X_j, X_i \cap X_t = \emptyset \quad (i \neq t)$.

В случае, когда класс F таков, что множества $\{x: f(x) \leq 0\}$ - многомерные параллелепипеды, нахождение μ - подсистем системы (1) может быть осуществлено с помощью методов теории линейных неравенств. Один из таких методов для таксономии реализован в пакете «КВАЗАР».

§ 2. Метод потенциальных функций

Здесь кратко изложим только геометрический смысл метода. Пусть $|X| \subset R^n$ - конечное множество, которое надо подвергнуть таксономии. Положим $f(x) = \sum_{y \in X} K(x, y)$.

Если множество X распадается на таксоны x_1, \dots, x_k , то график функции $f(x)$ представляет собой «горный ландшафт» – вершины над множествами x_1, \dots, x_k , а между множествами – «ущелья», разделяющие таксоны.

§ 3. Обзор некоторых методов таксономии

3.1. Метод сфер

Изложим один из алгоритмов разработанных и программно реализованных Н. Г. Загоруйко и другими.

Заданы: конечное множество $|X| \subset R^n$ и радиус r гиперсферы. Строим шар $B(x, r)$, где x - любая точка из X , затем множество $X \cap B(x, r)$. Пусть z_1 - центр тяжести множества $X \cap B(x, r)$. Строим $X \cap B(z_1, r)$, определяем z_2 как центр тяжести этого множества. Строим $X \cap B(z_2, r)$, находим центр тяжести z_3 этого множества и т.д.

Этот процесс заканчивается, когда $z_{i+1} = z_i$. Полагаем в качестве первого таксона $X_1 = X \cap B(z_1, r)$. Затем повторяем всё изложенное для множества X/X_1 , получаем таксон X_2 и т.д.

3.2. Метод «масок»

Пусть задано конечное множество X векторов $x = (\xi_1, \dots, \xi_n) \in R^n$ с координатами $\xi_i \in \{0, 1\}$. Разбиваем множество X на таксоны X_i по «сходству» элементов, т.е. по числу совпадающих координат. В качестве порога выберем некоторое число $\alpha > 0$. Выбираем любой элемент $x_1 \in X$ и запоминаем как элемент таксона X_1 . Выбираем произвольно следующий элемент x из X и сравниваем с x_1 . Если число $r = r(x, x_1)$ совпадающих координат не меньше, чем α , то векторы считаем «похожими» и выбранный элемент x отбрасываем. Если же $r < \alpha$, то полагаем элемент $x = x_2$, и вектор x_2 запоминаем как элемент таксона X_2 и т.д. Таким образом выделяются «маски» x_1, \dots, x_k . Далее для всех элементов множества X формируются «знаки». «Знак» вектора x есть k -

мерный вектор, j -я координата которого равна 1, если $r(x, x_j) \geq \alpha$, и равна 0 в противном случае.

Все векторы с одним знаком попадают в один таксон x_j , этот знак – «имя» таксона.

3.3. Прочие методы

Выше приведены наиболее простые методы таксономии. В целом, многообразие методов достаточно велико, и многие из этих методов связаны с весьма сложными критериями оценки качества разбиения на таксоны. В книге Н. В. Загоруйко [11] подробно изучается сфера методов таксономии. В частности в ней описывается метод, в котором реализованы критерии, выявленные в результате исследования требований, предъявленных людьми – экспертами к качеству разбиения на таксоны. Среди этих требований близость точек друг к другу внутри таксона, удаленность таксонов друг от друга, одинаковость структуры расположения точек внутри таксона, приблизительная одинаковость числа точек в каждом таксоне.

Рассмотрим ещё некоторые примеры методов таксономии.

1. Метод, использующий линейные неравенства, для разделения на два таксона. Пусть $X = \{x_1, \dots, x_m\}$. Требуется найти $J \subset \overline{1, m}$ такое, чтобы система $\langle x_j, x \rangle > 0 \ (\forall j \in J), \langle x_j, x \rangle < 0 \ (\forall j \notin J)$ была совместна, и чтобы $|J| \geq \alpha \cdot m, \alpha \in [0, 1]$.

2. Метод «объединения» (Э. М. Браверманна). Пусть $X = \{x_1, \dots, x_m\}$. Пусть введена некоторая мера $K(A, B)$ близости между множествами A и B . Пусть x_1, \dots, x_k - некоторое начальное разбиение множества X на таксоны. Надо получить разбиение на ℓ множеств, $\ell > k$. На первом шаге из k

подмножеств получаем $k-1$, объединяя в одно те два множества X_i и X_j , для которых $K(X_i, X_j) = \max_{p=q} K(X_p, X_q)$.

Глава 6.

МЕТОДЫ ВЫБОРА ПРИЗНАКОВОГО ПРОСТРАНСТВА

§ 1. Оценка признаков по коэффициентам разделяющей функции

Существует интуитивно понятное соображение, что если классификация проводится в соответствии с линейной дискриминантной функцией от вектора $x = (\xi_1, \dots, \xi_n)$:

$$f(x) = \sum_{i=1}^n \alpha_i \xi_i,$$

то величина $|\alpha_i|$ является некоторой оценкой информативности i -го признака.

Именно, пусть два образа имеет вид:

$$X_1 = \{a \in R^n : \|a\| = 1, \langle a, x^* \rangle \geq 0\};$$

$$X_2 = \{b \in R^n : \|b\| = 1, \langle b, x^* \rangle \leq 0\},$$

$$\text{где } \|x^*\| = 1.$$

Напомним обозначения: если $x = (\xi_1, \dots, \xi_n)$, $J = \{i_1, \dots, i_k\} \subset \overline{1, n}$, то $x(J) = (\xi_{i_1}, \dots, \xi_{i_k})$, $X_1(J) = \{x(J) : x \in X_1\}$.

Положим $\mathcal{P}(J) = X_1(J) \cap X_2(J)$.

Объем $\nu(\mathcal{P}(J))$ этого множества можно считать мерой неразделимости множеств X_1 и X_2 по системе J признаков.

Приведем без доказательства теорему В. В. Андреева:

Теорема. Если $x^* = (\xi_1^*, \dots, \xi_n^*)$ и $c = (\alpha_1, \dots, \alpha_n)$, то

$$\nu(\mathcal{P}(J)) = \nu\left(\left\{c \in R^J : F_J(c) \leq 0\right\}\right), \text{ где } F_J(c) = \left(\sum_{i \in J} \alpha_i \xi_i^*\right)^2 - \left(1 - \sum_J \alpha_i^2\right) \left(1 - \sum_J \xi_i^{*2}\right).$$

Следствие. $v(P(J)) = v_i \cdot \left(1 - \sum_j \xi_i^{*2}\right)^{1/2}$, где v_i - объем шара единичного радиуса в R' .

Согласно этому следствию естественно ввести такую оценку информативности подсистемы J признаков:

$$\omega(J) = 1 - \left(1 - \sum_j \xi_i^{*2}\right)^{1/2}.$$

§ 2. Прочие методы

Пусть $\omega(J)$ - некоторая мера информативности подсистемы J признаков, $J \subset \overline{1, n} = I$. Если $J = \{i_1, \dots, i_s\}$ $\omega(J) = \omega(i_1) + \dots + \omega(i_s)$, то задача

$$\max \{ \omega(J) : |J| \leq p, J \subset \overline{1, n} \} \quad (1)$$

решается очень просто: нужно только найти наибольшие по информативности p признаков. Если, однако, признаки не являются независимыми, то для решения задачи (1) в общем случае приходится просматривать C_n^p сочетаний признаков. При больших p и n с этим связан большой объем вычислений.

Поэтому применяются методы частичного перебора, не гарантирующие нахождения абсолютного оптимума.

Один из таких методов: метод Мерилла и Грина: исключаем признак i_1 , дающий $\max_i \omega(I \setminus \{i\})$, затем из $I \setminus \{i_1\}$ исключаем по тому же принципу следующий признак и т.д. Количество шагов равно $n + (n-1) + \dots + (n-p) \ll C_n^p$.

Метод Барабаша: находим признак i_1 , на котором достигается $\max_i \omega(i)$. Затем находим i_2 , реализующее $\max_{i \neq i_1} \omega(\{i_1, i\})$ и т.д. Число шагов $n + (n-1) + \dots + (n-p)$.

Метод случайного поиска с адаптацией (Лбов) состоит в следующем. Вначале признакам присваиваются одинаковые, равные $1/n$, вероятности их употребления. В соответствии с этими вероятностями выбираются несколько подсистем, каждая из которых содержит p признаков. Для каждой подсистемы J подсчитывается значение $\omega(J)$ её информативности. Пусть для этих подсистем:

$$\omega(J_1) = \max_J \omega(J), \quad \omega(J_2) = \min_J \omega(J).$$

Тогда для признаков входящих в J_1 , увеличиваем вероятность из использования, а для признаков из J_2 вероятности уменьшаем. Затем вся процедура повторяется для принятых новых значений вероятностей использования признаков.

Метод тупиковых тестов (Ю. И. Журавлев) также позволяет получить некоторые оценки информативности признаков: более информативным считается признак, вошедший в большее число тупиковых тестов.

Опишем ещё один метод, основанный на использовании принципе двойственности в линейном программировании.

Выделим некоторые подмножества признаков: $J_1 \subset \overline{1, n}, \dots, J_m \subset \overline{1, n}$. Пусть материал обучения составляет некоторое множество $S \subset R^n$. Предположим, что при использовании вместо S множества $S(J_i)$ доля правильных классификаций равна γ_i . Введем обозначение:

$$a_{ji} = \begin{cases} 1, & j \in J_i; \\ 0, & j \notin J_i. \end{cases}$$

Составим следующую таблицу:

Признак	Вариант			Ресурс (макс. частота употребления признаков)
	J_1	...	J_m	
1	a_{11}	...	a_{1m}	1
...
n	a_{1n}		a_{nm}	1
Доля правильных ответов	γ_1		γ_m	
Частота употребления вариантов	ξ_1		ξ_m	

Сформулируем задачу P линейного программирования:

$$\max \left\{ \sum_{i=1}^n \gamma_i \xi_i : \sum_{i=1}^m a_{ji} \xi_i \leq 1 \quad (\forall j \in \overline{1, m}), \xi_i \geq 0 \right\};$$

Она имеет смысл максимизации ожидания доли правильных ответов при смешанной стратегии использования вариантов подсистем признаков. Двойственная задача P^* имеет вид:

$$\min \left\{ \sum_{j=1}^m \eta_j : \sum_{j=1}^m a_{ji} \eta_j \geq \gamma_i \quad (\forall i \in \overline{1, n}), \eta_j \geq 0 \right\}.$$

Координаты её оптимального вектора $\bar{y} = (\bar{\eta}_1, \dots, \bar{\eta}_m)$ естественным образом интерпретируются как оценки дефицитности (ценности) признаков. Следовательно, для любой подсистемы $J \subset \overline{1, n}$ признаков можно измерять её информативность величиной $\omega(J) = \sum_{i \in J} \bar{\eta}_i$.

Глава 7.

ПРИКЛАДНЫЕ ЗАДАЧИ РАСПОЗНАВАНИЯ ОБРАЗОВ

Задачи распознавания образов мы будем рассматривать с точки зрения процессов автоматизации получения и обработки данных, анализа закономерностей в массивах информации и структуризации этих массивов.

С этой точки зрения задачи распознавания образов могут быть разбиты на следующие группы:

1. Выделение, формирование и оценка существенных признаков (факторов); преобразование пространства признаков.

2. Анализ закономерностей размещения элементов конечного множества в многомерном фазовом пространстве.

3. Разбиение пространства на классы эквивалентности, отнесение элементов пространства к классам.

В соответствии с этим в сфере приложений методов распознавания образов к решению практических задач выделим следующие области:

I. Оценка ситуаций.

В этой области можно выделить задачи интерпретации обстановки, ориентации, узнавания, диагностики.

II. Структуризация информации, организация материала наблюдений.

Здесь отметим три группы задач:

а. Задачи классификации, разбиение совокупности объектов, задачи районирования.

в. Запоминание информации, заполнение пробелов в ней, обобщение опыта, его сжатие и упорядочение, агрегирование.

с. Создание типологии; выделение типопредставителей, эталонов.

III. Моделирование.

Выделим следующие группы задач:

а. Обратные задачи; идентификация моделей.

в. Построение разделяющих поверхностей, изоповерхностей; моделирование эмпирических зависимостей, целей и предпочтений.

с. Задачи поиска и выбора, прогнозирования, управления, планирования, проектирования.

Далее будут приведены более или менее конкретные примеры приложений.

§ 1. Применение методов распознавания образов для учёта плохо формализуемых ограничений в алгоритмах оптимального планирования

Рассмотрим задачу математического программирования, интерпретируемую как задачу проектирования сложного объекта. В этой модели объект описывается его вектором состояния: $x = (\xi_1, \dots, \xi_m) \in R^n$, где ξ_i - значение i - параметра состояния. Пусть показателями качества функционирования объекта в состоянии x будут значения функций: $f_0(x), f_1(x), \dots, f_m(x)$.

Предположим, что для нормального функционирования объекта необходимо, чтобы

$$\alpha_j \leq f_j(x) \leq \beta_j (j = 1, \dots, m). \quad (1)$$

Кроме того, выбор состояния x подчинён требованию максимизации $f_0(x)$. Получаем задачу математического программирования

$$\max \{f_0(x) : (1)\}. \quad (2)$$

Сформулированную модель весьма широко и успешно применяют в решении практических задач, в том числе задач оптимального технико-экономического планирования и проектирования. Однако её применение в ряде случаев наталкивается на значительные трудности. Одна из таких трудностей состоит в том, что некоторые из ограничений (1) могут быть плохо формализуемыми. В принятой нами интерпретации это может означать следующее. Специалисты по проектированию объектов рассматриваемого класса на основании своего опыта могут выделить ряд условий, которым должны удовлетворять параметры ξ_1, \dots, ξ_n , но не могут записать эти условия аналитически, т.е. вид некоторых функций f_j в системе (1) им не известен.

Однако по некоторым интуитивным соображениям, исходя из опыта, эти специалисты (будем далее называть их экспертами) для многих векторов состояния $\bar{x} = (\bar{\xi}_1, \dots, \bar{\xi}_n)$ могут судить о их допустимости по j -му ограничению, т.е. о выполнении или невыполнении неравенства

$$\alpha_j \leq f_j(\bar{x}) \leq \beta_j.$$

Моделирование плохо формализуемых ограничений при этих условиях может быть осуществлено с помощью методов дискриминантного анализа. Пусть следующее ограничение плохо формализуемо:

$$\alpha_i \leq f_i(\bar{x}) \leq \beta_i \quad (3)$$

Предположим, что на основании опыта экспертам известны: конечное множество $M_i \subset R^n$, Состоящее из некоторых векторов x , удовлетворяющих условию (3), и конечное множество $N_i \subset R^n$, содержащее векторы, не удовлетворяющие этому условию.

Тогда можно решить следующую задачу дискриминантного анализа: найти функцию $g(x)$ из некоторого класса функций G , разделяющую множества M_i и N_i , т.е.

$$g(x) > 0 \quad (\forall x \in M_i), \quad g(x) < 0 \quad (\forall x \in N_i), \quad g \in G.$$

Пусть $g_i(x)$ - решение этой системы. Тогда в задаче (2) вместо ограничения (3), неизвестного нам как аналитическое выражение, записываем ограничение $g_i(x) > 0$.

Пусть, \bar{x} – оптимальный вектор получившейся задачи оптимизации. Вектор \bar{x} предъявляется для экспертизы, которая устанавливает его допустимость или недопустимость по ограничению (3). Если \bar{x} удовлетворяет этому ограничению, то расширяем множество M_i , добавляя к нему вектор \bar{x} ; в противном случае \bar{x} добавляем к N_i . Затем решаем задачу дискриминации для расширения множеств M_i и N_i , и так далее.

§ 2. Задачи медицинской диагностики и управления процессом лечения

Предположим, что вектор $x=(\xi_1, \dots, \xi_n)$ есть вектор параметров, характеризующих состояние человека (пульс, давление крови, температура; качественные признаки, кодируемые числами, например, наличие или отсутствие головной боли и т.п.). Пусть $y=(\eta_1, \dots, \eta_m)$ - вектор состояния окружающей среды.

С помощью методов дискриминантного анализа можно осуществить классификацию (диагностику), ставя в соответствие $(m+n)$ - мерному вектору $(x,y)=(\xi_1, \dots, \xi_n, \eta_1, \dots, \eta_m)$ класс людей, страдающих соответствующим заболеванием.

Одним из классов может быть класс здоровых людей. Обучение правилу классификации проводят на основе материала обучения, включающего из каждого класса некоторое конечное множество его представителей.

За этапом диагностики следует этап выбора метода лечения. Выбираемая процедура лечения характеризуется своим вектором параметров $z=(\zeta_1, \dots, \zeta_p)$. Характеристики состояния человека после лечения зависят от векторов x, y, z . Обозначим эти характеристики через

$$f_0(x, y, z), f_1(x, y, z), \dots, f_q(x, y, z).$$

Каждая характеристика должна находиться в некоторых допустимых пределах: $\alpha_j \leq f_j(x, y, z) \leq \beta_j$ ($j = 0, 1, 2, \dots, q$)

При этих ограничениях и при фиксированных векторах $x = \bar{x}$, $y = \bar{y}$ может быть поставлена задача отыскания процедуры – такой, что ей соответствует максимально возможное значение показателя $f_0(\bar{x}, \bar{y}, z)$ --- .

Однако некоторые из зависимостей f_j могут плохо формализуемыми. В этом случае с помощью методов

дискриминантного анализа для них могут быть построены приближённые модели, как об этом сказано в § 1.

Список литературы

1. Мазуров Вл. Д. О построении комитета системы выпуклых неравенств. – Кибернетика, 1967. – №2.
2. Мазуров Вл. Д. О некоторых математических конструкциях для распознавания образов // Нелинейное программирование и приложения. – Свердловск, 1979.
3. Мазуров Вл. Д., Казанцев В. С., Белецкий Н. Г., Мезенцев С. В., Сачков Н. О. Пакет КВАЗАР прикладных программ распознавания образов. – Свердловск, 1979.
4. Нильсон Н. Обучающие машины. – М., 1968.
5. Черников С. Н. Свёртывание конечных систем линейных неравенств. – ДАН УССР, 1969. – № 1.
6. Яненко Н. Н. Проблемы математической технологии. – Тезисы доклада международной конференции «Структура и организация пакетов программ». – Тбилиси, 1976.
7. Журавлёв Ю. И. Об алгебраическом подходе к решению задач распознавания или классификации. – Проблемы кибернетики. – 1978. – № 3.
8. Айзерман М. А., Браверманн Э. М., Розоноэр Л. И. Метод потенциальных функций в теории обучения машин. – М., 1970.
9. Горелик А. Л., Скрипкин В. А. Построение систем распознавания. – М., 1974.
10. Бонгард М. М. Проблема узнавания. – М., 1967.
11. Загоруйко Н. Г. Методы распознавания и их применение. – М., 1972.
12. Метод комитетов в распознавании образов // Труды Института математики и механики УНЦ АН СССР. – Свердловск, 1974. – Вып.6.

ОГЛАВЛЕНИЕ

Введение	3
Глава 1. Основные математические модели распознавания образов	5
§ 1. Содержательный смысл задачи распознавания образов	5
§ 2. Моделирование объекта классификации	9
§ 3. Модель дискриминантного анализа	13
§ 4. Модель таксономии	19
§ 5. Выбор признакового пространства	22
Глава 2. Линейный дискриминантный анализ	25
§ 1. Методы дискриминантного анализа	25
§ 2. Разделяющие возможности аффинных функций	30
§ 3. Метод линейной коррекции	35
§ 4. Методы линейного программирования	39
Глава 3. Метод комитетов в дискриминантном анализе	45
§ 1. Комитетные конструкции	45
§ 2. Теоремы существования	51
§ 3. Алгоритмы построения комитетов	59
§ 4. Построение минимального комитета	63
§ 5. Вопросы программного обеспечения	71
Глава 4. Обзор методов дискриминантного анализа	73
§ 1. Метод потенциальных функций	73
§ 2. Методы, основанные на теории статистических решений	76
§ 3. Методы алгебры логики	79
Глава 5. Методы таксономии	83
§ 1. Метод выделения максимальных совместных подсистем	83

§ 2. Метод потенциальных функций	84
§ 3. Обзор некоторых методов таксономии	85
Глава 6. Методы выбора признакового пространства	88
§ 1. Оценка признаков по коэффициентам разделяющей функции	88
§ 2. Прочие методы	90
Глава 7. Прикладные задачи распознавания образов	93
§ 1. Применение методов распознавания образов для учёта плохо формализуемых ограничений в алгоритмах оптимального планирования	94
§ 2. Задачи медицинской диагностики и управления процессом лечения	97
Список литературы	99

Мазуров Владимир Данилович

Математические методы распознавания образов

Учебное пособие

Редактор Н. Алексеева

Подписано в печать 10.09.2010

Формат 60х34 1/16

Уч.изд.л. 7,8

Усл.печ.л. 6,4

Тираж 100 экз.

Заказ 93
